

Màster Interuniversitari en Estadística i Investigació Operativa UPC-UB

Títol: Anàlisi de dades procedents de telefonia mòbil per la modelització de la demanda de transport

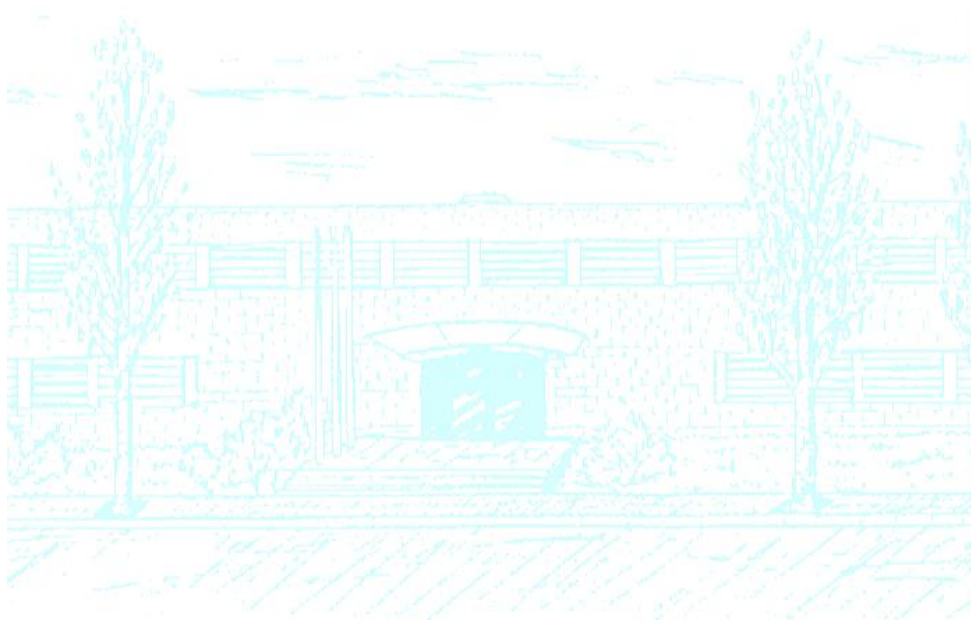
Autor: Francesc Gasparín Casajust

Directora: Lúdia Montero Mercadé

Departament: Estadística i Investigació Operativa

Universitat: Universitat Politècnica de Catalunya

Convocatòria: Juny 2018



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



Màster en Estadística i Investigació Operativa

**Anàlisi de dades procedents de telefonia mòbil
per la modelització de la demanda de transport**

Francesc Gasparín Casajust

Directora: Lúdia Montero Mercadé

Departament d'Estadística i Investigació Operativa

Agraïments

Aquest TFM (Treball final de màster) no hagués estat possible sense l'ajuda de la Lúdia Montero, per això m'agradaria agrair que m'hagi guiat i aconsellat durant tot el procés del treball. La seva implicació, col·laboració i el continu seguiment han sigut essencials per al correcte desenvolupament d'aquest treball.

També agrair a la meua família el suport i la motivació que m'han transmès durant tot l'estudi.

Gràcies.

Resum

Paraules clau: Models de Transport UTPS, demanda de transport, matrius OD, dades TIC, data science

MSC2000: 62-07

Els models de transport són eines necessàries per a la planificació del transport, sobretot a les ciutats mitjanes i grans. També permeten fer prediccions sobre futurs escenaris de sistemes de transport. Dins de les metodologies estàndard més utilitzades hi ha el model clàssic de quatre etapes UTPS (Urban Transportation Planning System). El model clàssic, en concret per al Projecte Virtual Mobility LAB, requereix matrius OD (origen-destinació) entre zones de transport, obtingudes mitjançant les noves tecnologies (mètodes actuals), concretament a partir de la telefonia mòbil. Aquestes matrius han estat elaborades per l'empresa Kineo per al projecte Virtual Mobility LAB. Virtual Mobility LAB és un projecte que desenvolupa un model UTPS per esdevenir un marc comú per donar suport en la presa de decisions sobre el disseny i l'avaluació dels nous conceptes de mobilitat a l'Àrea Metropolitana de Barcelona, principalment a la Primera Corona de Barcelona. Per l'anàlisi i validació de les matrius OD procedents de telefonia mòbil, s'estudien les relacions entre les variables significatives pel sector per a poder trobar patrons de mobilitat, consistents amb els coneixements i informes públics sobre la demanda del transport en la Primera Corona Metropolitana. A més, s'apliquen nous mètodes estadístics (per primer cop en aquest sector), com l'Anàlisi de Components Principals (ACP) i l'Anàlisi de Correspondències Simples (ACS) per sintetitzar els patrons característics espai-temporals dels perfils de mobilitat.

Abstract

Keywords: UTPS Transportation Models, transport demand, OD matrix, ICT data, data science

MSC2000: 62-07

Transport models are necessary tools for transport planning, especially in medium and large cities. They also make predictions about future scenarios of transport systems. Within the most commonly used standard methodologies is the classic four-stage model UTPS (Urban Transportation Planning System). The classic model, specifically for the Virtual Mobility LAB Project, requires OD (origin-destination) matrices between transport zones, obtained through new technologies (current methods), in particular from a mobile phone. These matrices have been prepared by the Kineo company for the Virtual Mobility LAB project. Virtual Mobility LAB is a project that develops an UTPS model to become a common framework to support decision-making about the design and evaluation of the new concepts of mobility in the Metropolitan Area of Barcelona, principally in the First Crown of Barcelona. For the analysis and validation of OD arrays from mobile phone, the relationships between the significant variables in the sector are studied to be able to find patterns of mobility, consistent with the knowledge and public reports on the demand for transport in the First Crown Metropolitan. In addition, new statistical methods are applied (for the first time in this sector), such as the Principal Component Analysis (PCA) and the Correspondence Analysis (CA) to synthesize the spatial-temporal characteristics patterns of mobility profiles.

Índex

Introducció.....	1
1. Modelització del transport	3
1.1. Introducció	3
1.2. Característiques dels problemes del transport.....	3
1.2.1. Característiques de la demanda del transport	4
1.2.2. Característiques de l'oferta del transport	4
1.3. La estructura clàssica del model de transport.....	5
1.3.1. Generació de viatges	7
1.3.2. Modelització de la distribució de viatges.....	8
1.3.3. Repartiment modal.....	13
1.3.4. Assignació.....	20
2. Mètodes de recollida de dades de transport.....	26
2.1. Mètodes tradicionals pels patrons de mobilitat	26
2.2. Recollida de dades de sensors	27
2.3. Mètodes actuals	28
2.3.1. Dades telefonia mòbil	29
2.3.2. Dades de seguiment de flotes de taxis	30
2.3.3. Dades de targetes intel·ligents de trànsit	31
2.3.4. Dades de xarxes socials	32
3. Dades de demanda de Kineo	33
3.1. Introducció	33
3.2. Dades.....	33
3.2.1. Matriu de dades	34
4. Projecte Virtual Mobility LAB.....	36
4.1. Un enfocament metodològic per construir el nucli de simulació del “Virtual Mobility Lab”.....	37
4.2. Model de la Primera Corona de l'Àrea Metropolitana de Barcelona.....	37
5. Metodologia.....	40
5.1. Anàlisi de Components Principals (ACP).....	41

5.2. Clustering	44
5.2.1. K-means clustering	45
5.3. Anàlisi de Correspondències Simples (ACS).....	46
6. Descripció del tractament de les dades	51
6.1. Variables de tipus numèric	51
6.2. Variables categòriques	52
6.2.1. Selecció de les variables	57
7. Resultats	58
7.1. Preprocessament de les dades	58
7.2. Dades de mobilitat a grans àmbits a la Regió Metropolitana de Barcelona (RMB)	58
7.3. Dades de mobilitat a la Primera Corona.....	64
7.4. Distribució horària dels desplaçaments a la RMB	72
7.5. Mapa de calor	74
7.6. Anàlisi de Components Principals (ACP).....	76
7.7. Classificació a partir de l'ACP	80
7.8. Anàlisi de Correspondències Simples (ACS).....	81
8. Conclusions	88
Bibliografia.....	92
Apèndix	94
A. Matriu de viatges origen-destinació	95
B. Franja horària per municipi origen.....	96
C. Franja horària per municipi destinació.....	97
D. Franja horària per districte origen de Barcelona	98
E. Franja horària per districte destinació de Barcelona	99
F. Codi R	100

Llista de figures

Figura 1.1: El clàssic model de transport en quatre etapes.....	6
Figura 1.2: Produccions i atraccions de viatges	7
Figura 2.1: Exemple de la xarxa cel·lular	29
Figura 4.1: Diagrama del flux de treball del “Procés de construcció de models dirigits a dades”	37
Figura 4.2: Candidats a l'àrea d'orientació del model Visum per al “Virtual Mobility LAB”	38
Figura 5.1: Construcció d'una taula de contingència.....	46
Figura 5.2: Núvol de perfils fila a l'ACS.	49
Figura 5.3: Núvol de perfils columna a l'ACS.....	49
Figura 7.1: Àrea de mobilitat RMB.....	59
Figura 7.2: Àrea de mobilitat segons el motiu del viatge RMB.....	60
Figura 7.3: Àrea de mobilitat segons el lloc de residència RMB	61
Figura 7.4: Mode de transport RMB	61
Figura 7.5: Mode de transport segons motiu del viatge RMB	63
Figura 7.6: Mode de transport segons lloc de residència RMB	63
Figura 7.7: Àrea de mobilitat PC.....	64
Figura 7.8: Àrea de mobilitat segons el motiu del viatge PC.....	65
Figura 7.9: Àrea de mobilitat segons el lloc de residència PC	66
Figura 7.10: Àrea de mobilitat segons el mode de transport PC	67
Figura 7.11: Mode de transport segons motiu PC	68
Figura 7.12: Àrea de mobilitat transport motoritzat PC	69
Figura 7.13: Àrea de mobilitat transport motoritzat PC	70
Figura 7.14: Franja horària RMB	72
Figura 7.15: Franja horària general i segons motiu, segons l'EMEF.....	72
Figura 7.16: Franja horària segons el motiu del viatge RMB	73
Figura 7.17: Franja horària segons el lloc de residència RMB	74
Figura 7.18: Mapa de calor OD municipis	74
Figura 7.19: Mapa de calor OD Municipis - Cotxe.....	75
Figura 7.20: Mapa de calor OD municipis – Autobús	75
Figura 7.21: Mapa de calor OD municipis - Metro.....	75
Figura 7.22: Projecció 1-2 dels individus.....	76
Figura 7.23: Projecció 1- 2 dels individus.....	77
Figura 7.24: <i>Cos2</i> de les variables numèriques als eixos 1-2	78
Figura 7.25: Projecció de les variables numèriques als eixos 1-2.....	78

Figura 7.26: <i>Cos2</i> de les variables numèriques als eixos 3-4	79
Figura 7.27: Projecció de les variables numèriques als eixos 3-4.....	79
Figura 7.28: Projecció variable municipi origen eix 1-2.....	80
Figura 7.29: Projecció variable municipi destinació eix 1-2.....	80
Figura 7.30: ACS municipis OD	82
Figura 7.31: ACS municipis origen.....	82
Figura 7.32: Viatges entre municipis OD.....	83
Figura 7.33: Viatges entre municipis, franja horària 7-10h.....	84
Figura 7.34: Viatges entre municipis, franja horària 13-15h.....	84
Figura 7.35: Viatges entre municipis, franja horària 17-20h.....	84
Figura 7.36: Viatges entre els barris de Barcelona.....	85
Figura 7.37: Mapa barris de Barcelona.	85
Figura 7.38: Viatges entre municipis origen de la PC i districtes destinació Barcelona.....	86
Figura 7.39: Mapa districtes de Barcelona.	86

Llista de taules

Taula 7.1: Àrea de mobilitat RMB	59
Taula 7.2: Àrea de mobilitat segons el motiu del viatge RMB	59
Taula 7.3: Àrea de mobilitat segons el motiu del viatge RMB (%)	60
Taula 7.4: Àrea de mobilitat segons el lloc de residència RMB	60
Taula 7.5: Àrea de mobilitat segons el lloc de residència RMB (%)	61
Taula 7.6: Mode de transport RMB	61
Taula 7.7: Mode de transport segons motiu del viatge RMB	62
Taula 7.8: Mode de transport segons motiu del viatge RMB (%)	62
Taula 7.9: Mode de transport segons lloc de residència RMB	63
Taula 7.10: Mode de transport segons lloc de residència RMB (%)	63
Taula 7.11: Àrea de mobilitat PC	64
Taula 7.12: Àrea de mobilitat segons el motiu del viatge PC	65
Taula 7.13: Àrea de mobilitat segons el motiu del viatge PC (%)	65
Taula 7.14: Àrea de mobilitat segons el lloc de residència PC	66
Taula 7.15: Àrea de mobilitat segons el lloc de residència PC (%)	66
Taula 7.16: Àrea de mobilitat segons el mode de transport PC.....	66
Taula 7.17: Àrea de mobilitat segons el mode de transport PC (%).....	67
Taula 7.18: Mode de transport segons motiu PC.....	68
Taula 7.19: Mode de transport segons motiu PC (%).....	68
Taula 7.20: Àrea de mobilitat transport motoritzat PC.....	69
Taula 7.21: Àrea de mobilitat transport motoritzat PC (%).....	69
Taula 7.22: Transport motoritzat segons el motiu del viatge PC	70
Taula 7.23: Transport motoritzat segons el motiu del viatge PC (%)	70
Taula 7.24: Ràtio transport públic/privat.....	71
Taula 7.25: Clúster ACP.....	81

Introducció

Els problemes de mobilitat i transport són i seran un greu problema en els països desenvolupats i en vies de desenvolupament, a causa del creixement econòmic i de la població que fan augmentar la flota de vehicles i la demanda del transport. Per estudiar els sistemes de transport es necessiten eines de la modelització del transport, on es segueix un formalisme ben establert i consensuat, anomenat el model clàssic de transport en quatre etapes (Ortuzar and Willumsen 2011) o UTPS (Urban Transportation Planning System) (Plummer 1962). Aquestes etapes són: generació de viatges, modelització de la distribució de viatges, repartiment modal i assignació.

El model clàssic de transport en quatre etapes és el marc de referència per la planificació del transport. Cadascuna de les etapes s'expliquen en aquesta memòria, des de la perspectiva del Projecte Virtual Mobility Lab (VML). El Projecte Virtual Mobility LAB (VML) és una plataforma integrada de modelització multimodal dirigida a ajudar en el disseny i avaluació d'impactes de nous conceptes de mobilitat i transport. L'àmbit d'estudi és la Primera Corona de Barcelona, però relacions amb la resta de l'Àrea Metropolitana de Barcelona ('*gates*' del model) també hi són presents.

La primera etapa (generació de viatges) té com a objectiu modelar el nombre total de viatges generats per un origen i atrets per una destinació, a cada zona d'estudi (TAZ, Transport Analysis Zone). La segona etapa és la modelització de la distribució de viatges, aquí es representa el patró de mobilitat en forma de matriu de viatges o matriu origen-destinació (matriu OD). En aquesta etapa, es destaquen els mètodes del factor de creixement i models de gravetat. La tercera etapa (repartiment modal) segurament és la més important en la planificació del transport, i està adreçada a la modelització de l'elecció de mode de transport. Els models representatius són: models sintètics (repartiment modal agregats) i models d'elecció discreta (lògit multinomial, entre d'altres). Per finalitzar, s'arriba a l'etapa d'assignació, es tracta de la confrontació oferta-demanda, de vehicles i/o persones (demanda) a les xarxes de transport (oferta). Cal destacar que el transport privat i públic no té el mateix procés d'assignació, ja que aquest darrer és més complicat de tractar.

Un altre apartat molt important és la forma de recollida de dades en tot el procés de modelització UTPS (Plummer 1962). En el cas concret de l'obtenció de dades de la

demanda de transport, es troben els mètodes tradicionals (enquestes), complementades per dades procedents de sensors de trànsit i els mètodes actuals (relacionats amb les noves tecnologies). Les dades que s'utilitzen en aquest treball són provinents dels mètodes actuals, concretament de la telefonia mòbil. En el Capítol 2, s'explica detalladament com són aquestes dades.

Les dades concretes que s'utilitzen han estat elaborades per l'empresa Kineo pel Projecte VML. Kineo és una empresa tecnològica que analitza i visualitza dades de geolocalització per diverses finalitats, en molts casos, alimenta models adreçats a la presa de decisions. El format de les dades elaborades per Kineo s'explica detalladament en el Capítol 3 i els mètodes utilitzats per analitzar-les i validar la seva consistència s'inclou en el Capítol 5.

A l'hora de realitzar l'anàlisi de les matrius OD s'han tingut en compte les possibles relacions, per exemple: modes de transport i àrea territorial, lloc de residència i àrea territorial, franges horàries, entre altres relacions que s'han considerat rellevants per l'estudi dels patrons de mobilitat. Els resultats obtinguts es comparen amb els resultats de l'Enquesta de Mobilitat en Dia Feiner 2015 (EMEF 2015). Seguidament, s'ha aplicat l'Anàlisi de Components Principals (ACP) per visualitzar la consistència temporal de les dades de mobilitat OD horàries, variables numèriques i representar de forma clara i visual les variacions, relacions, oposicions i altres conceptes. Un cop analitzats els perfils horaris, s'ha aplicat un procés de segmentació de les relacions OD, a partir de l'ACP, concretament s'utilitza el k-means. La finalitat d'aplicar aquesta tècnica és trobar patrons o grups (clústers) dins d'un conjunt de parelles OD, on aquestes estan dins d'un mateix grup, són similars entre elles i diferents de les d'altres grups. Per acabar, s'utilitza l'Anàlisi de Correspondències Simples (ACS) per validar la consistència espacial de les dades de mobilitat OD i trobar relacions i oposicions entre grups de viatges totals/modals relacionats amb variables categòriques referides a àrees territorials (municipi, barri, districte, etc.).

Per finalitzar, s'extrauen conclusions sobre els resultats obtinguts a partir de l'anàlisi estadística, i es proposen diferents opcions per aprofundir l'estudi dels patrons de mobilitat procedents tant de fonts de dades tradicionals, com de noves tecnologies.

Per acabar la introducció, voldria explicar la motivació per triar aquest tema com a Treball Final de Màster. Estic interessat en temes relacionats amb el transport públic, l'elecció de la millor ruta i teoria de cues, entre d'altres. El punt de vista que presenta el treball és diferent del qual s'està acostumat. Sembla interessant aprofundir sobre les tècniques per l'anàlisi dels patrons de mobilitat a partir d'unes dades de telefonia mòbil, que en un primer moment hom no pensa que permetin realitzar un estudi de patrons de mobilitat.

1. Modelització del transport

1.1. Introducció

Un model (Ortuzar and Willumsen 2011) és una representació del món real, que es vol centrar en elements específics, importants des d'un cert punt de vista. Poden haver-n'hi tant de físics, com d'abstractes. Els primers, entre altres, són aquells que s'utilitzen en arquitectura o en mecànica de fluids dirigits al disseny (models mentals). Els models mentals tenen un paper molt important en la comprensió i interpretació del món real, i es milloren a través de discussions, entrenament i l'experiència.

Dins dels models abstractes, una de les classes més importants són els models matemàtics. Aquests intenten simular o copiar el món real mitjançant equacions matemàtiques basades en la teoria sobre el camp concret d'aplicació. Encara que s'ha avançat molt en l'ús d'aquests models, ja que són representacions simplifiades, poden ser molt complexos, requereixen gran quantitat de dades i són de gran valor per discutir i valorar polítiques tàctiques o estratègiques. A part d'això, un gran avantatge que tenen és que durant la seva formulació, calibratge i ús, es pot aprendre molt del sistema que representen, i així s'enriqueixen els models mentals permetent així una gestió més intel·ligent del sistema tractat.

1.2. Característiques dels problemes del transport

Al llarg de la història els problemes de mobilitat i transport han estat un greu problema en els països avançats i en augment, a causa del creixement econòmic i de la població, a més han fet créixer la demanda del transport i el trànsit provocant congestió, retard, accidents i contaminació.

Aquests problemes no desapareixeran en un futur. Tot això és a causa d'una planificació de transport molt deficient al llarg del temps. Actualment, s'està en disposició de prendre decisions acurades i orientades a maximitzar els avantatges de les transformacions i introducció de noves formes de transport, minimitzant costos

conòmics, operatius, tàctics i estratègics i efectes secundaris negatius en el medi ambient i les persones.

1.2.1. Característiques de la demanda del transport

La demanda de transport es produeix per satisfer una necessitat de mobilitat deguda a activitats produïdes en llocs concrets, per exemple, els estudiants han de desplaçar-se a la Universitat per assistir a les seves classes (activitats). Per comprendre-la s'ha d'estudiar i d'entendre la distribució d'aquestes activitats a l'espai, depenen del seu context.

La demanda té moltes característiques específiques: hora del dia, dia de la setmana, motiu del viatge, etc. Per tal de donar servei a les necessitats de mobilitat, les autoritats desenvolupen serveis. Els serveis poden ser infraestructures per facilitar la mobilitat en transport privat o disseny d'una xarxa de línies de transport públic. Si un servei no té els atributs que el client demana serà inútil. Resulta molt difícil analitzar i predir la demanda de transport.

La demanda de transport té lloc a l'espai. El tractament de l'espai més comú és dividir-lo en àrees d'estudi. Aquestes s'anomenen zones de transport o TAZ (Transport Analysis Zones) i es codifiquen amb les xarxes de transport corresponents.

Finalment, la demanda i l'oferta del transport tenen elements dinàmics molt forts. La major part de la demanda es concentra en unes hores del dia, on s'hi produeix la congestió, entre altres problemes. El sistema de transport no està preparat per donar servei a una demanda màxima, ja que està ajustat per a la mitjana diària. Per això, les variacions màximes i mínimes de la demanda continuaran essent un problema central en la modelització i la planificació del transport.

Si un sistema de transport públic es dimensionés exclusivament per la situació de demanda màxima, aleshores en hores vall estaria infrautilitzat amb tot l'extra cost que això suposaria. Les puntes de demanda s'han de tractar específicament.

1.2.2. Característiques de l'oferta del transport

La principal característica de l'oferta de transport és que és un servei i no un bé. O sigui, no es pot emmagatzemar. Per tant, s'ha de consumir quan es té, en cas contrari, es perd. Per això, és molt important estimar la demanda amb la precisió més gran com sigui possible.

Un dels problemes més grans és que la infraestructura i els vehicles no són propietat ni operen per la mateixa empresa. Aquest és el cas de la majoria de modes de transport. Això genera un greu problema, ja que la separació de poders entre el proveïdor d'infraestructures i el proveïdor del servei de transport genera unes interaccions bastant complexes.

També existeix un problema econòmic, ja que construir aquestes infraestructures té un cost molt elevat i es tarda anys a implantar-les. És necessària una provisió elevada, ja que és important des del punt de vista de l'oferta. A causa d'això, aquestes infraestructures s'han utilitzar al complet, no es pot usar la meitat d'un aeroport per exemple. A més, la inversió en transport té un gran paper polític. Es necessita un gran estudi per construir les infraestructures per veure si hi ha zones afectades, i més si la construcció es realitza a superpoblacions.

Hi ha diversos problemes en l'oferta: accidents, contaminació, degradació del medi ambient, congestió, entre d'altres. En alguns d'ells rarament els usuaris assumiran costos d'aquests efectes. Per tant, si s'internalitzessin aquests costos, es podria ajuda a prendre millors decisions i millorar l'assignació de la demanda a modes alternatius.

La congestió és una de les característiques més importants de l'oferta de transport. És una paraula difícil de definir, donat que a tots els llocs no es considera per igual. La congestió és la condició d'un flux vehicular que es veu saturat a causa de l'excés de demanda de les vies, tant en l'àmbit urbà com interurbà, produint increments en els temps de viatge i embussos. D'altra banda, un dels inconvenients més importants que provoca la congestió és una greu contaminació atmosfèrica. Aquest problema és més comú en els països en desenvolupament que no pas als desenvolupats, això pot ser per diversos factors, incloent-hi la manca de manteniment del vehicle i l'ús insuficient dels sistemes de control d'emissions de vehicles.

La contaminació atmosfèrica i els seus impactes en la salut pública generen una gran preocupació per part de la comunitat de recerca en salut ambiental, les agències reguladores ambientals, les indústries i el públic. La qualitat de l'aire, tant a l'interior com a l'exterior, està estretament relacionada amb la morbiditat i la mortalitat per malalties respiratòries i cardiovasculars. Els contaminants aeris més comuns provocats pel tràfic són els que inclouen partícules (*PM*), ozó (*O₃*), monòxid de carboni (*CO*), diòxid de sofre (*SO₂*), diòxid de nitrogen (*NO₂*), plom (*Pb*), compostos orgànics volàtils (*VOC*) i hidrocarburs aromàtics policíclics (*HAP*).

1.3. La estructura clàssica del model de transport

Els anys d'experimentació i desenvolupament han donat lloc a una estructura general que s'ha anomenat model de transport clàssic. Aquesta estructura és, en efecte, un resultat de la pràctica en la dècada de 1960 (*UTPS Urban Transport Planning System model paradigm*), però s'ha mantingut més o menys inalterada malgrat les grans millores en les tècniques de modelatge des de llavors.

La generalització del model està representada a la Figura 1.1. Primerament es considera un sistema de zones de transport (Zonificació) i xarxes viàries i de transport públic que caracteritzen l'oferta de transport de l'àmbit, i la recollida i codificació de

dades de planificació, calibratge i validació. Aquesta base de dades és una base anual per a cada zona de l'àrea d'estudi. Aquestes dades s'utilitzen per estimar un model del nombre total de viatges generats i atrets per cada zona de l'àrea d'estudi (generació de viatges). El següent pas és la seva distribució a l'espai, produint així una matriu de mobilitat entre zones d'origen i destinació dels viatges (matriu OD). La següent etapa és la modelització de l'elecció del mode de transport, o sigui, construcció de la matriu OD per als diferents modes. Finalment, l'última etapa del model clàssic requereix l'assignació per cada modalitat a les seves xarxes corresponents: confrontació oferta-demanda.

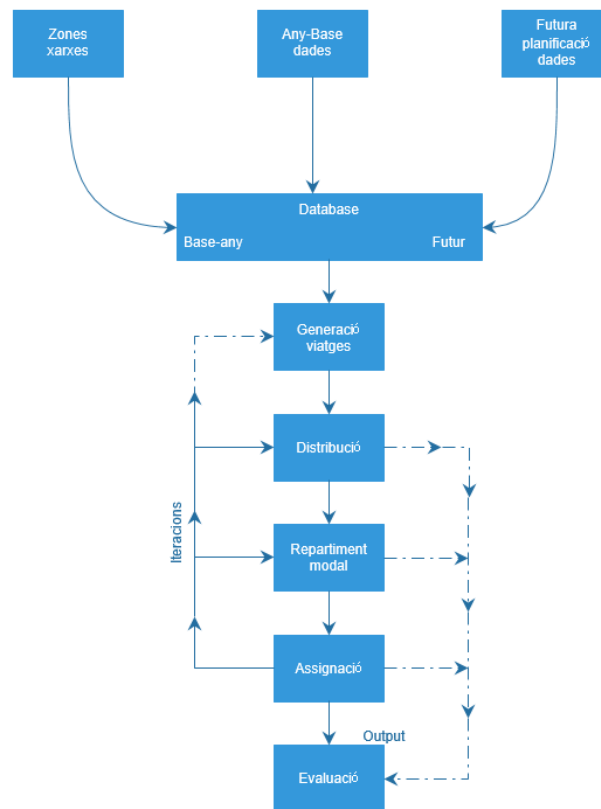


Figura 1.1: El clàssic model de transport en quatre etapes

Els procediments d'assignació són diferents en el transport privat i en el públic. En el transport privat consisteix a calcular quins són els camins emprats pels desplaçaments entre parelles origen-destinació (parelles OD) i quin flux de vehicles/viatgers carrega cada camí.

Els procediments d'assignació de transport públic són més complexes, en general, i determinen les estratègies òptimes d'ús de les línies dels diferents modes existents, intercanviadors i línies de transport entre parelles OD.

El model clàssic es presenta com una seqüència de quatre submodels: generació de viatges, distribució, repartiment modal (modal *split*) i assignació. Generalment, les decisions que es prenen durant els viatges no s'adopten en aquest tipus de seqüències.

1.3.1. Generació de viatges

L'etapa de generació de viatges del model de transport clàssic té com a objectiu modelar el nombre total de viatges generats per un origen i atrets per una destinació, a cada zona d'estudi (TAZ).

A continuació, es defineixen alguns conceptes bàsics i s'examinen alguns dels factors que afecten la generació i l'atracció de viatges.

Primerament, es realitzen diferents qüestions sobre els tipus de viatges: quin és l'esdeveniment d'interès, activitats amb un període curt de temps, desplaçament d'un lloc a un altre, o una seqüència d'aquests viatges, entre d'altres. El modelador ha de pensar amb totes aquestes opcions per crear un conjunt de la següent llista per a desenvolupar i aplicar un conjunt de models.

- Trajecte o viatge: es tracta d'un moviment d'un sol sentit des d'un origen fins a una destinació.
- Viatge a casa: és on l'origen del viatger és l'origen o la destinació del viatge.
- Viatge fora de casa: és aquell on el final del viatge no és la llar del viatger.
- Producció de viatges: es defineix com a final d'origen d'un viatge a casa o com a origen d'un viatge fora de casa (Figura 1.2).
- Atracció de viatge: es defineix com a final no casolà d'un viatge a casa o la destinació d'un viatge fora de casa (Figura 1.2).
- Generació de viatges: es defineix com el nombre total de viatges generats en una zona.
- Període de residència: curt període d'estada en un lloc determinat.
- Activitat: interès associat a un motiu, però no necessàriament a un lloc fix.
- Viatge o cadena de viatges: un conjunt d'activitats i viatges vinculats.

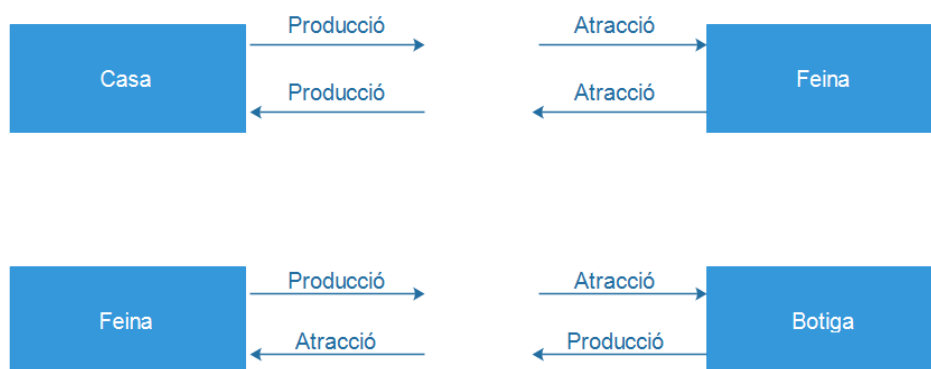


Figura 1.2: Produccions i atraccions de viatges

També hi ha diferents agrupacions de la caracterització dels viatges, que són:

- Motiu
- Hora del dia
- Tipus de persona

Finalment, els factors que afecten la generació de viatges són:

- Produccions de viatges personals
- Atraccions de viatges personals
- Produccions de viatges de mercaderies i atraccions

Un cop es té identificada la llista, les agrupacions i els factors es procedeix a la modelització de la generació de viatges mitjançant regressió lineal, classificació creuada o anàlisi de categories i enfocament de la categoria de la persona.

Un exemple per resumir les diferents agrupacions i factors de la generació de viatges és: la quantitat de viatges per motiu laboral d'una zona residencial està relacionada amb el nombre d'habitatges de la zona i el grau d'envelliment de la seva població.

1.3.2. Modelització de la distribució de viatges

Com s'ha vist fins ara, les produccions i els llocs d'interès proporcionen una idea del volum de realització de viatges en una àrea d'estudi, però això no és suficient per a la modelització i la presa de decisions. El que es necessita és una millor idea del patró de la realització de viatges o patró de mobilitat, és a dir, des d'on i fins a on es fan els viatges, a banda dels modes de transport triats i les rutes preses.

El patró de mobilitat es pot representar, almenys amb dues maneres diferents. El primer és com una 'matriu de viatges' o 'taula de viatges'. Això registra els viatges realitzats d'un origen a una destinació durant un període de temps determinat; també s'anomena matriu origen-destinació (O-D) i pot ser desglossada per tipus de persona (grup socioeconòmic, per exemple definit pel gènere i el grup d'edat) i el motiu de l'activitat realitzada al final de cada viatge. Aquesta representació en forma de matriu és necessària per a tots els models d'assignació.

La segona forma de presentar un patró de mobilitat és considerar els factors que generen i atreuen viatges, és a dir, una base de producció-atracció (P-A), en general, la casa és considerada com la finalitat "productora", i el treball, la botiga, etc., com "atraient" final. Per necessitat, una matriu P-A cobrirà un període de temps més llarg (normalment un dia) que una matriu O-D.

S'han proposat diversos mètodes al llarg dels anys per modelar la distribució dels viatges (submodel de generació de viatges) entre destinacions; algunes de les més simples només són adequades per a estudis tàctics a curt termini, on no es preveuen canvis importants en l'accessibilitat proporcionada per la xarxa. Altres semblen respondre millor als canvis en el cost de la xarxa i, per tant, es suggereixen estudis estratègics a més llarg termini o per a aquells tàctics que impliquin canvis importants en els preus de transport relatius; sovint es basen en P-A.

La distribució del viatge sovint es considera un problema agregat. Tanmateix, l'elecció de la destinació també es pot tractar com un problema d'elecció (desglossat) discreta i tractat amb models desagregats a nivell de l'individu.

En aquesta etapa del model clàssic s'introdueixen mètodes que només responen a taxes de creixement relatives en els orígens i les destinacions; són adequats per a l'extrapolació de tendència a curt termini. També es pot utilitzar una família de models sintètics, el més conegut és el model de gravetat. Un aspecte important de l'ús de models sintètics és el seu calibratge, és a dir, la tasca de fixar els seus paràmetres perquè el patró de viatge de base sigui ben representat pel model. S'han proposat també altres models sintètics i el més important d'ells, és el model d'intervenció.

A continuació, s'explicarà breument dos mètodes: mètodes del factor de creixement, concretament el doblement restringit, i la distribució del model de gravetat o sintètic. Aquests són els més importants relacionats amb aquest treball. Un tractament detallat es pot trobar al Capítol 5 (Ortuzar and Willumsen 2011).

1.3.2.1. Mètodes del factor de creixement

Els mètodes del factor de creixement són uns mètodes que només responen a les taxes de creixement relatives en els orígens i les destinacions i això és adequat per a l'extrapolació de tendència a curt termini. Primerament es donen unes definicions i notacions, que serveixen per introduir i poder començar a explicar els factors de creixement doblement restringits i la distribució del model de gravetat.

Definicions i notacions

El patró de mobilitat es representa en una àrea d'estudi mitjançant una matriu de viatges O-D. Es tracta bàsicament d'una matriu bidimensional de cel·les, on les files i columnes representen cadascuna de les zones de l'àrea d'estudi.

Les cel·les de cada fila contenen els viatges originats en aquesta zona que tenen com a destinacions les zones de les columnes corresponents. La diagonal principal correspon a desplaçaments intrazonals. Per tant: T_{ij} és el nombre de viatges entre origen i i destinació j ; la matriu total és $\{T_{ij}\}$ o T ; O_i és el nombre total de viatges provinents de la zona i ; i D_j és el nombre total de viatges atrets per la zona j . P_i és la quantitat de viatges produïts o generats en una zona i i Q_j els atrets per la zona j .

S'utilitzen les minúscules, t_{ij} , o_i i d_j per indicar observacions d'una mostra o d'un estudi anterior; i les majúscules representen l'objectiu o els valors que s'intenten modelar per al període de modelatge corresponent.

Les matrius es poden desagregar encara més, per exemple, per tipus de persona (n) i/o per mode/motiu (k). Per tant:

- T_{ij}^{kn} són viatges de i a j pel motiu k i el tipus de persona n ;
- O_i^{kn} és el nombre total de viatges provinents de zona i pel motiu k i el tipus de persona n , i així successivament.

La suma dels viatges en una fila hauria de ser igual al nombre total de viatges que provenen d'aquesta zona; la suma dels viatges en una columna hauria de correspondre a la quantitat de viatges atrets per aquesta zona. Aquestes condicions es poden escriure com:

$$\sum_j T_{ij} = O_i \quad (1.1)$$

$$\sum_i T_{ij} = D_j \quad (1.2)$$

Es pot considerar l'element de costos en termes de distància, temps o unitats monetàries. Sovint és convenient utilitzar una mesura que combini tots els atributs principals relacionats amb el cost d'un viatge (C_{ij}) i això normalment es coneix com el cost generalitzat del viatge o impedància del viatge. Normalment, aquesta és una funció lineal dels atributs del viatge ponderats per coeficients que intenten representar la seva importància relativa tal que el viatger el percebi. Una possible representació d'aquest per al motiu k pot ser (suprimint superíndex k):

$$C_{ij} = a_1 t_{ij}^v + a_2 t_{ij}^w + a_3 t_{ij}^t + a_4 t_{ij}^n + a_5 F_{ij} + a_6 \phi_j + \delta \quad (1.3)$$

on el significat dels atributs del viatge són:

- t_{ij}^v és el temps de viatge del vehicle entre i i j ;
- t_{ij}^w és el temps de caminar de o des de la parada (estació) o des de la zona d'estacionament;
- t_{ij}^t és el temps d'espera a les parades (o el temps gastat buscant lloc per aparcar);
- t_{ij}^n és el temps d'intercanvi, si n'hi ha;
- F_{ij} és un càrrec monetari: la tarifa que es carrega per viatjar entre i i j o el cost d'utilitzar el cotxe per a aquest viatge, incloent-hi els peatges o càrrecs de congestió (s'ha de tenir en compte que sovint els costos operatius dels cotxes no es perceben bé i que els mitjans electrònics de pagament tendeixen a difuminar d'alguna manera el vincle entre l'ús i el pagament);
- ϕ_j és un cost de terminal (normalment d'estacionament i és aplicable al transport privat) associat amb el viatge de i a j ;
- δ és una penalització modal, un paràmetre que representa tots els altres atributs no inclosos en la mesura generalitzada fins ara;
- $a_{1...6}$ són pesos units a cada element de cost; tenen dimensions adequades per a la conversió de tots els atributs a unitats comunes.

Factors de creixement restringits doblement

Quan es té informació disponible sobre la quantitat futura de viatges que s'originen i que finalitzen a cada zona es genera un problema interessant. Això implica diferents taxes de creixement per als viatges dins i fora de cada zona i , en conseqüència, tenen dos conjunts de factors de creixement per a cada zona, per exemple, τ_i i Γ_j . L'aplicació d'un factor de creixement mitjà, per exemple $F_{ij} = 0.5(\tau_i + \Gamma_j)$, només és un compromís deficient que no es compliria cap dels dos objectius o les restriccions de finalització de viatge. Durant els anys s'han proposat diversos mètodes iteratius per obtenir una matriu de viatge estimada que sé que satisfà els objectius.

Tots aquests mètodes inclouen el càlcul d'un conjunt de coeficients de correcció intermedis que després s'apliquen a les entrades de cel·les en cada fila o columna, segons correspongui. Després d'aplicar aquestes correccions per fila, els totals de cada columna es calculen i es comparen amb els valors objectiu. Si les diferències són significatives, els nous coeficients de correcció es calculen i s'apliquen quan sigui necessari.

En el transport, aquests mètodes són coneguts pels seus autors com Fratar als Estats Units i Furness en altres llocs. Per exemple, Furness (1965) va introduir els "factors d'equilibri" A_i i B_j de la manera següent:

$$T_{ij} = t_{ij} \cdot \tau_i \cdot \Gamma_j \cdot A_i \cdot B_j \quad (1.4)$$

- o incorporant les taxes de creixement en noves variables a_i i b_j :

$$T_{ij} = t_{ij} \cdot a_i \cdot b_j \quad (1.5)$$

- on $a_i = \tau_i \cdot A_i$ i $b_j = \Gamma_j B_j$.

Els factors a_i i b_j (o A_i i B_j) s'han de calcular de manera que compleixin les restriccions (1.1) i (1.2). Això s'aconsegueix en un procés iteratiu que es resumeix a continuació:

1. Establir tots els $b_j = 1.0$ i resoldre per a_i ; significa trobar els factors de correcció a_i que satisfan les restriccions de generació de viatges.
2. Amb l'últim a_i que és resolt per b_j , per exemple, satisfer les restriccions d'atracció de viatge.
3. Mantenint fixat b_j , resoldre per a_i i repetir els passos (2) i (3) fins que els canvis siguin prou petits.

Aquest mètode produeix solucions dins del 3 al 5% dels valors objectiu en algunes iteracions quan es compleixen determinades condicions. Un grau de convergència més estricte pot ser important des de la perspectiva de la consistència del sistema del model. Aquest mètode es denominava sovint un "algorisme bi-proporcional" a causa de la naturalesa de les correccions que comporta. El problema no es limita al transport, també a: trànsit telefònic i actualització de matrius d'entrada i sortida en economia.

Però en qualsevol cas, el mètode Furness intenta produir les correccions mínimes a la matriu d'anys de base necessàries per satisfer les restriccions de viatges de futur.

La condició més important requerida per a la convergència d'aquest mètode és que les taxes de creixement produeixen valors objectiu T_i i T_j tals que:

$$\sum_i \tau_i \sum_j t_{ij} = \sum_j \Gamma_j \sum_i t_{ij} = T \quad (1.6)$$

L'execució d'aquesta condició pot requerir la correcció de les estimacions de viatges finals produïdes pels models de generació de viatge.

Distribució del model de gravetat

S'han desenvolupat models de distribució d'un altre tipus per a la prognosi de patrons de mobilitat futurs quan es produeixin canvis importants a la xarxa. Es comencen a partir de suposicions sobre el comportament del viatge en grup i la influència de factors externs, com ara els desplaçaments totals i la distància recorreguda. El més conegut d'aquests models és el model de gravetat, originalment generat a partir d'una analogia amb la llei gravitacional de Newton. Aquests models estimen els desplaçaments per a cada cel·la de la matriu sense utilitzar directament el patró de viatge observat en una mostra o una situació obsoleta; per tant, de vegades es denominen models sintètics a diferència dels models de creixement.

Probablement, el primer ús rigorós d'un model de gravetat va ser de Casey (1955), qui va suggerir un enfocament d'aquest tipus per sintetitzar els viatges de compra i les àrees de captació entre les ciutats d'una regió. En la seva formulació més senzilla, el model té la següent forma funcional:

$$T_{ij} = \frac{\alpha P_i P_j}{d_{ij}^2} \quad (1.7)$$

- on P_i i P_j són les poblacions de les ciutats d'origen i de destinació, d_{ij} és la distància entre i i j , α és un factor de proporcionalitat (amb unitats de *viatges · distància²/població²*).

Es va considerar aviat una analogia massa simplista amb la llei gravitacional i les primeres millores incloïen l'ús de totals de viatges (O_i i D_j) en lloc de poblacions totals, i un paràmetre n per al calibratge com a potència per a d_{ij} . Aquest nou paràmetre no es va limitar a ser un nombre enter i diferents estudis van estimar els valors entre 0.6 i 3.5.

El model es va generalitzar assumint que l'efecte de distància o "separació" es podria modelar millor per una funció decreixent, que cal especificar, de la distància o del cost generalitzat de viatge entre les zones. Això es pot escriure com:

$$T_{ij} = \alpha O_i D_j f(c_{ij}) \quad (1.8)$$

- on $f(c_{ij})$ és una funció generalitzada dels costos de viatge amb un o més paràmetres de calibratge. Aquesta funció sovint rep el nom de "funció de dissuasió" perquè representa l'element dissuasiu de viatjar a mesura que augmenta la distància (el temps) o el cost generalitzat. Les versions populars d'aquesta funció són:

$$f(c_{ij}) = \exp(-\beta c_{ij}) \text{ funció exponencial} \quad (1.9)$$

$$f(c_{ij}) = c_{ij}^{-n} \text{ funció de potència} \quad (1.10)$$

$$f(c_{ij}) = c_{ij}^n \exp(-\beta c_{ij}) \text{ funció combinada} \quad (1.11)$$

En aquest projecte, es proposa la formulació i l'estimació de models de distribució sintètics per motius.

1.3.3. Repartiment modal

L'elecció del mode de transport és probablement una de les fases del model clàssic més important en la planificació del transport. Això es deu al paper clau que juga el transport públic en la formulació de polítiques. Gairebé sense excepció viatjar en els modes de transport públic, és utilitzar l'espai viari amb més eficiència i produeix menys accidents i emissions que l'ús d'un automòbil privat.

La qüestió de l'elecció de mode, per tant, és probablement l'element més important en la planificació del transport i la formulació de polítiques. Afecta l'eficiència general amb què es pot viatjar a les zones urbanes, la quantitat d'espai urbà dedicat a les funcions de transport i si hi ha una varietat d'opcions disponibles per als viatgers. El tema també és important en el transport interurbà, ja que, de nou, els modes ferroviaris poden proporcionar un mode de transport més eficient (en termes de recursos consumits, inclòs l'espai), però també hi ha una tendència a augmentar els viatges per carretera.

És important llavors desenvolupar i utilitzar models que siguin sensibles a aquells atributs del viatge que influeixen en les opcions individuals de mode. Es veurà fins a quin punt es pot satisfer aquesta necessitat utilitzant enfocaments agregats, on polítiques alternatives s'han d'expressar com a modificacions de funcions d'utilitat, o sigui inflexibles, com ara el cost generalitzat dels viatges.

Els models que s'explicaran amb profunditat, ja que, són els més adients en aquest treball són: models sintètics (repartiment modal) i models d'elecció discreta, concretament el model lògit multinomial.

1.3.3.1. Distribució i estructures del repartiment modal

Primerament, es defineixen les propietats que té el cost compost (K_{ij}^n) de viatjar entre i i j que percep el tipus de persona n :

- $K \leq \text{Min}_k \{C^k\}$
- $\lim_{\lambda \rightarrow \infty} K = \text{Min}_k \{C_k\}$, aquesta és l'opció de mode "tot o res"
- $\frac{dK}{dC^k} = P^k$

La primera d'aquestes propietats significa que quan s'afegeix una nova alternativa, encara que sigui molt poc atractiva en principi, els costos compostos es reduiran o, com a màxim, es mantindran igual. La segona propietat posa de manifest la importància de λ_n com un pes associat a l'elecció segons els costos generalitzats i està relacionat amb la variabilitat dels costos generalitzats dels diferents modes. Per a un λ_n molt gran, el model predirà una opció "tot o res" de l'alternativa de menor cost generalitzat.

El model es troba sovint a la pràctica en aplicacions agregades, en particular en àrees urbanes. Tanmateix, un dels problemes a la pràctica és que els modeladors de vegades no comproven si es compleix la restricció, però és molt possible que la restricció no estigui satisfeta. Si aquest és el cas, els models combinats (gravetat i elecció de mode) produiran resultats patològics.

S'ha trobat a la pràctica que la mateixa estructura pot ser diferent per als diferents motius de viatge. Normalment, l'estructura correcta seria G/D/MS/A per al viatge al treball i G/MS/D/A per a altres finalitats. Això reflectiria una condició on és més fàcil canviar la destinació, per exemple, un viatge de compres que no pas el canvi de mode.

S'ha de tenir en compte el cas de l'estructura G/MS/D/A, on es comença calculant el cost compost pel mode n d'arribar a totes les destinacions de cada origen i :

$$K_i^n = \frac{-1}{\beta} \log \left(\sum_j \exp(-\beta C_{ij}^n) \right) \quad (1.12)$$

A continuació, els costos compostos s'utilitzen per obtenir les divisions de mode per origen i , en el cas d'elecció binària entre les alternatives 1 i 2:

$$P_i^1 = \frac{1}{1 + \exp(-\lambda(K_i^2 - K_i^1))} \quad (1.13)$$

Els models de gravetat separats es desenvolupen utilitzant els costos de cada modalitat amb les generacions de viatge de manera específica, però compartint els totals d'atracció de viatges. Tot i que es tracta d'un model d'elecció de mode basat en la generació, es tenen en compte els costos d'arribar a cada destinació per cada modalitat.

Per a moltes aplicacions, aquests models agregats segueixen sent vàlids i en ús. Tanmateix, per a un maneig més refinat de les característiques i preferències personals,

s'han proposat models desagregats que responen millor als elements clau en l'elecció de mode i fan un ús més eficient dels esforços de recopilació de dades.

1.3.3.2. Models discrets d'elecció

Es comença per algunes consideracions generals i es passa a explicar el marc teòric, la teoria d'utilitat aleatòria, en què es presenten aquests models. Això serveix per introduir una terminologia bàsica i presentar la dualitat del model individual que és tan útil per comprendre el que la teoria postula. A continuació, es presenta un cas particular, el del model lògit multinomial, que proporciona un conjunt d'eines de modelatge potent i molt emprat.

Consideracions generals

Els models de transport de demanda agregats, es basen en les relacions observades per a grups de viatgers o en les relacions mitjanes en nivell de zona de transport (TAZ). D'altra banda, els models de desagregació de la demanda es basen en les opcions observades pels viatgers individuals o les famílies. Aquest marc permet desenvolupar models més realistes.

Per representar l'atractiu de les alternatives, s'utilitza el concepte d'utilitat (que és un marc teòric definit com el que l'individu busca maximitzar). Les alternatives, de per si, no produeixen utilitat: això es deriva de les seves característiques i les de l'individu; per exemple, la utilitat observable sol definir-se com una combinació lineal de variables, com ara:

$$V_{car} = 0.25 - 1.2 \cdot IVT - 2.5 \cdot ACC - 0.3 \cdot \frac{C}{I} + 1.1 \cdot NCAR \quad (1.14)$$

on cada variable representa un atribut del viatger. La influència relativa de cada atribut, en termes de contribuir a la satisfacció general produïda per l'alternativa, ve donada pel seu coeficient. Per exemple, un canvi d'unitat en el temps d'accés (ACC) té aproximadament el doble de l'impacte d'un canvi unitari en el temps de viatge en vehicle (IVT) i més de set vegades l'impacte d'un canvi d'unitat en el cost variable/ingressos (C/I). Les variables també poden representar característiques de l'individu; per exemple, s'espera que una persona que pertanyi a una casa amb un gran nombre d'automòbils (NCAR), tindria més probabilitats de triar l'opció de cotxe que una altra pertanyent a una família amb només un vehicle. La constant específica alternativa, 0.25 en l'equació, normalment s'interpreta com a representant de la influència neta de totes les característiques no observades o no incloses explícitament de l'individu i l'opció en la seva funció d'utilitat.

Per predir si es tria una alternativa, segons el model, el valor de la seva utilitat s'ha de contrastar amb els de les opcions alternatives i transformar-se en un valor de probabilitat entre 0 i 1. Per a això existeixen una varietat de transformacions matemàtiques que

típicament es caracteritzen per tenir una trama de S-forma, com ara, pel cas d'elecció binària entre 2 modes de transport:

- Lògit $P_1 = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)}$

Els models d'elecció discreta no es poden calibrar, en general, utilitzant tècniques estàndard d'ajust de corbes, com els quadrats mínims, perquè la seva variable dependent P_i és una probabilitat no observada (entre 0 i 1) i les observacions són les opcions individuals (que són 0 o 1); les úniques excepcions són models per a grups homogenis d'individus, o quan el comportament de cada individu es registra en diverses ocasions, perquè les freqüències d'elecció observades també són variables entre 0 i 1.

Algunes propietats útils d'aquests models van ser resumides per Spear (1977):

1. Els models desagregats de demanda (DM) es basen en les teories del comportament individual i no constitueixen analogies físiques de cap tipus. Per tant, com un intent d'explicar el comportament individual, un avantatge potencial important sobre els models agregats és que és més probable que els models DM siguin estables (o transferibles) en temps i espai.
2. Els models DM es calculen utilitzant dades individuals i això té les següents implicacions:
 - Els models DM poden ser més eficients que els models agregats en termes d'ús de la informació; es requereixen menys punts de dades, ja que cada opció individual s'utilitza com a observació. En la modelització agregada, una observació és la mitjana de centenars d'observacions individuals.
 - A mesura que s'utilitzen dades individuals, es pot utilitzar tota la variabilitat inherent a la informació.
 - Els models DM es poden aplicar, en principi, a qualsevol nivell d'agregació; però, tot i que això sembla obvi, els processos d'agregació no són trivials.
 - Els models DM són menys propensos a patir els biaixos a causa de la correlació entre unitats agregades. Un problema greu en utilitzar informació agregada és que el comportament individual pot estar amagat per les característiques no identificades associades a les zones; això es coneix com a correlació ecològica. Aquest fenomen pot passar, per exemple, si les característiques d'ús de la terra de la zona B són propícies per a més viatges a peu.
3. Els models desagregats són probabilístics; a més, donen la probabilitat d'escollir cada alternativa i no indiquen quina és la que s'ha seleccionat, l'ús ha d'estar fet de conceptes de probabilitat bàsics com ara:
 - El nombre esperat de persones que utilitzen una determinada opció de viatge és igual a la suma de cada una de les probabilitats d'escollir aquesta alternativa:

$$N_i = \sum_n P_{in}$$

- Un conjunt independent de decisions pot modelar-se per separat tenint en compte cada un com a elecció condicional; llavors les probabilitats resultants es poden multiplicar per produir probabilitats conjuntes per al conjunt, com ara:

$$P(f, d, m, r) = P(f)P\left(\frac{d}{f}\right)P\left(\frac{m}{d}, f\right)P\left(\frac{r}{m}, d, f\right)$$

amb f = freqüència,; d = destinació; m = mode; r = ruta

4. Les variables explicatives incloses en el model poden tenir coeficients estimats explícitament. En principi, la funció d'utilitat permet qualsevol nombre i especificació de les variables explicatives, a diferència del cas de la funció de cost generalitzada en models agregats que generalment es limita i té diversos paràmetres fixos. Això té implicacions com les següents:
 - Els models DM permeten una representació més flexible de les variables de política considerades rellevants per a l'estudi.
 - Els coeficients de les variables explicatives tenen una interpretació marginal d'utilitat directa (és a dir, reflecteixen la importància relativa de cada atribut).

Marc teòric

El marc teòric o paradigma més comú per generar models d'elecció discreta és la teoria d'utilitat aleatòria (Domencich i McFadden 1975; Williams 1977), que bàsicament postula que:

1. Les persones pertanyen a una determinada població homogènia Q , quan actuen racionalment i posseeixen informació perfecta, és a dir, sempre seleccionen aquesta opció que maximitza la seva utilitat personal neta (fins i tot l'espècie ha estat identificada com "*Homo economicus*") subjecte a requisits legals, socials, físics i /o restriccions pressupostàries (tant en temps com diners).
2. Hi ha un determinat conjunt $A = \{A_1, \dots, A_j, \dots, A_N\}$ d'alternatives disponibles i un conjunt X de vectors d'atributs mesurats dels individus i les seves alternatives. Un individu donat està dotat d'un conjunt particular d'atributs $x \in X$ i en general s'enfrontarà a un conjunt d'opcions $A(q) \in A$.

A partir d'aquí, s'assumirà que el conjunt d'elecció de l'individu està predeterminat; això implica que l'efecte de les restriccions ja s'ha tingut en compte i no afecta el procés de selecció entre les alternatives disponibles. Es considera la determinació de l'elecció, juntament amb altres qüestions pràctiques importants.

3. Cada opció $A_j \in A$ ha associat una utilitat neta U_{jq} per a l'individu q . El modelador, que és un observador del sistema, no posseeix informació completa sobre tots els elements considerats per l'elecció individual; per tant, el modelador suposa que U_{jq} pot estar representat per dos components:
 - Una part mesurable, sistemàtica o representativa V_{jq} que és una funció dels atributs mesurats x .

- Una part aleatòria ϵ_{jq} que reflecteix les idiosincràsies i els gustos particulars de cada individu, juntament amb qualsevol error de mesura o observació realitzat pel modelador.

Així, el modelador postula que:

$$U_{jq} = V_{jq} + \epsilon_{jq} \quad (1.15)$$

que permet explicar dues aparents "irracionalitats": que dues persones amb els mateixos atributs i que tinguin el mateix conjunt d'opcions poden seleccionar diferents opcions, i que alguns individus no sempre poden seleccionar el que sembla ser la millor alternativa (des del punt de vista dels atributs considerats pel modelador).

Perquè la descomposició sigui correcta, es necessita certa homogeneïtat en la població estudiada. En principi, s'exigeix que totes les persones comparteixin el mateix conjunt d'alternatives i s'enfrontin a les mateixes restriccions i, per aconseguir-ho, es necessita prèviament segmentar el mercat.

Tot i que s'ha anomenat V representant que porta el subíndex q perquè és una funció dels atributs x i això pot variar d'individual a individual. D'altra banda, sense pèrdua de generalitat es pot suposar que els residus ϵ són variables aleatòries amb una mitjana 0 i una determinada distribució de probabilitat que cal especificar. Una expressió popular i simple per a V és:

$$V_{jq} = \sum_k \theta_{kj} x_{jkq} \quad (1.16)$$

on s'assumeix que els paràmetres θ són constants per a tots els individus en el conjunt homogeni (model de coeficients fixos), però poden variar entre alternatives.

És important destacar l'existència de dos punts de vista en la formulació del problema anterior: en primer lloc, aquell de l'individu que pesa tranquil·lament tots els elements d'interès (sense cap aleatorietat) i selecciona l'opció més convenient; en segon lloc, la del modelista que, mitjançant l'observació de només alguns dels elements anteriors, necessita els residus ϵ per explicar el que d'altra manera suposaria un comportament no racional.

4. L'individu q selecciona l'alternativa de màxima utilitat, és a dir, l'individu tria A_j si i només si:

$$U_{jq} \geq U_{iq}, \forall A_i \in A(q) \quad (1.17)$$

això és

$$V_{jq} - V_{iq} \geq \epsilon_{iq} - \epsilon_{jq} \quad (1.18)$$

Com que l'analista ignora el valor de $(\epsilon_{iq} - \epsilon_{jq})$ no és possible determinar amb certesa si es manté. Així, la probabilitat de triar A_j està donada per:

$$P_{jq} = Prob\{\epsilon_{iq} \leq \epsilon_{jq} + (V_{jq} - V_{iq}), \forall A_i \in A(q)\} \quad (1.19)$$

i si es desconeix la distribució conjunta dels residus ϵ , no és possible derivar una expressió analítica del model. El que sí que es sap, és que els residus són variables aleatòries amb certa distribució que podem denotar per $f(\epsilon) = f(\epsilon_1, \dots, \epsilon_N)$. S'ha notat de passada que la distribució U , $f(U)$, és la mateixa però amb una mitjana diferent (és a dir, V en lloc de 0).

Per tant, es pot escriure (1.19) de manera més concisa com:

$$P_{jq} = \int_{R_N} f(\epsilon) d\epsilon \quad (1.20)$$

on

$$R_N = \begin{cases} \epsilon_{iq} \leq \epsilon_{jq} + (V_{jq} - V_{iq}), & \forall A_i \in A(q) \\ V_{jq} + \epsilon_{jq} \geq 0 \end{cases}$$

i models de diferents formes es poden generar depenent de la distribució dels residus ϵ .

Una classe important de models d'utilitat aleatoris és la generada per funcions d'utilitat amb residus independents i distribuïts de forma idèntica (IID). En aquest cas, $f(\epsilon)$ es pot descompondre en:

$$f(\epsilon_1, \dots, \epsilon_N) = \prod_n g(\epsilon_n)$$

on $g(\epsilon_n)$ és la distribució d'utilitat associada a l'opció A_n , i l'expressió general (1.20) es redueix a:

$$P_j = \int_{-\infty}^{\infty} g(\epsilon_j) d(\epsilon_j) \prod_{i \neq j} \int_{-\infty}^{V_j - V_i + \epsilon_j} g(\epsilon_i) d\epsilon_i \quad (1.21)$$

on s'ha ampliat el rang d'ambdós integrants a $-\infty$ (una lleugera incoherència) per resoldre'ls.

La interpretació geomètrica bidimensional d'aquest model, juntament amb les propostes del cas més general de correlació i variàncies desiguals, són presentades i discutides per Ortúzar i Williams (1982). L'equació (1.21) també es pot expressar com:

$$P_j = \int_{-\infty}^{\infty} g(\epsilon_j) d(\epsilon_j) \prod_{i \neq j} G(\epsilon_j + V_j - V_i) \quad (1.22)$$

amb

$$G(X) = \int_{-\infty}^X g(x) dx$$

i és interessant esmentar que s'ha dedicat una gran quantitat d'esforç a l'hora d'intentar trobar formularis adequats per a g que permetin (1.22) resoldre's en forma tancada.

Cal tenir en compte que el requisit de residus IID significa que les alternatives han de ser, de fet, independents.

Model lògit multinomial (MNL)

Aquest és el model d'elecció discreta a la pràctica més simple i popular. Es pot generar suposant que els residus aleatoris en (1.20) es distribueixen IID Gumbel (també anomenat valor extrem tipus I, EV1). Amb aquesta hipòtesi, les probabilitats d'elecció són:

$$P_{ig} = \frac{\exp(\beta V_{iq})}{\sum_{A_j \in A(q)} \exp(\beta V_{jq})} \quad (1.23)$$

on les funcions d'utilitat generalment tenen el lineal en la forma de paràmetres (1.16) i el paràmetre β (que es normalitza a un en la pràctica, ja que no es pot estimar separatament de la θ) està relacionat amb la desviació estàndard comuna de la variància EV1 mitjançant:

$$\beta = \frac{\pi}{\sigma\sqrt{6}} \quad (1.24)$$

El fet que β no pugui ser estimat separatament dels paràmetres θ en V_{iq} es coneix com a identificació teòrica; tots els models d'elecció discreta tenen problemes d'identificació, que requereixen establir determinats paràmetres en un valor determinat per estimar el model de forma exclusiva.

1.3.4. Assignació

En aquesta etapa es tracta l'assignació de vehicles i persones a les xarxes de transport públic i viari seguint un enfocament bastant intuïtiu per introduir algunes de les idees bàsiques rellevants.

1.3.4.1. Conceptes bàsics

S'adopta un enfocament formal que es centra en el postulat de l'equilibri d'usuari a la xarxa, o en nivell de sistema. El sistema de xarxa, i en el cas del transport públic, les característiques dels serveis oferts com la freqüència i la capacitat, representen els elements principals del costat de l'oferta de transport. Aquests són més o menys fixos a curt termini. Durant un període més llarg, les autoritats i operadors de transport canviaran tarifes, freqüències i tipus de vehicles; els administradors de xarxes de carreteres milloraran les carreteres existents (i construiran), limitaran l'estacionament i introduiran peatges i els càrrecs de congestió. Tot i que aquestes són representacions reals dels canvis de l'oferta que tenen impacte en l'augment de la demanda, no es tenen bons models per predir aquest tipus de canvis a llarg termini. Els models de xarxa són insuficients: només

són models de costos i com canvien els costos de transport amb diferents nivells de demanda. La tasca d'especificar un millor sistema de subministrament a més llarg termini correspon als responsables de la presa de decisions, als planificadors i als analistes.

En el pensament econòmic convencional, els intercanvis reals de béns i serveis es produeixen com a conseqüència de combinar la seva demanda amb l'oferta. El punt d'equilibri resultant d'aquesta combinació defineix el preu al qual s'intercanvien els béns i els seus respectius fluxos (quantitats intercanviades) al mercat. El punt d'equilibri es troba quan el cost marginal de produir i vendre els béns és igual a l'ingrés marginal obtingut de la venda. La teoria econòmica admet que aquest equilibri mai no pot passar realment a la pràctica, ja que el sistema de preus i nivells de producció es troba en un ajust permanent per afrontar els canvis en el poder adquisitiu, els gustos, la tecnologia i les tècniques de producció. Tanmateix, el concepte d'equilibri continua sent valuós per entendre el moviment de l'economia i pronosticar els seus estats futurs.

És útil considerar el sistema de transport en aquest context. Per la part de l'oferta (a curt termini), o més correctament el model de cost, es compon d'una xarxa de transport $S(L, C)$ representada pels arcs L (i els seus nodes associats) i els seus costos C . Els costos són funció d'un nombre d'atributs associats als arcs, p . I per la part de la demanda es compon d'una indicació del nombre de viatges per parella O-D i el mode que es faria per a un determinat desplaçament. Un dels principals elements que defineixen els nivells de servei és, en aquest context, el temps de viatge, però sovint els costos monetaris (tarifes, combustible) i característiques com la comoditat per al transport públic també poden ser rellevants. Si el nivell real de servei que ofereix la xarxa de transport resulta inferior al que es calcula, es podria esperar una reducció de la demanda i potser un canvi cap a altres destinacions, modes i/o hores del dia. La relació de flux de velocitat (o generada de fluxos de costos) és important, ja que relaciona l'ús de la xarxa amb el nivell de servei que pot oferir.

La xarxa de transport públic s'ha de definir en termes semblants a la xarxa privada. Tanmateix, hauria de contenir especificacions addicionals dels serveis oferts en funció de les seves rutes, capacitats, freqüència, tarifes i, idealment, encara que rares vegades a la pràctica, la seva qualitat, fiabilitat i regularitat.

En el cas d'un sistema de transport es pot veure l'equilibri que es produeix a diversos nivells. El més senzill és l'equilibri a la xarxa viària on els viatgers en mode privat (cotxe) en forma d'una matriu OD fixa cerquen rutes per minimitzar els costos del viatge (temps). Això dóna lloc a la cerca de rutes alternatives, explorant-ne de noves i, potser, estabilitzant-se en un patró relativament estable. Aquesta assignació de viatges a rutes produeix un patró de fluxos de viatges als camins i arcs que es pot dir que estan en equilibri quan els viatgers ja no poden trobar millors rutes cap a les seves destinacions: ja estan viatjant per les millors rutes disponibles donada la demanda de transport. Aquest és l'equilibri d'usuari de transport privat. Un fenomen similar, però potser menys dramàtic, es produeix a les xarxes de transport públic on els passatgers poden buscar estratègies (és

a dir, combinacions de serveis) per reduir els costos de viatge generalitzats que afecten els temps d'amuntegament, espera i caminada i els temps en els vehicles.

Hi ha, però, altres nivells d'interacció. A mesura que augmenta la congestió dels cotxes, els autobusos que operen a les mateixes carreteres també augmentaran els temps de viatge. Això pot induir a alguns usuaris de transport públic (i operadors d'autobusos) a canviar les seves estratègies per evitar retards. Aquestes opcions interactuen amb les dels conductors d'automòbil, ja que els nous arranjaments poden proporcionar un canvi en la capacitat en alguns arcs i, per tant, nous punts d'equilibri.

A un nivell superior, el patró de flux resultant pot afectar les opcions de mode, destinació i hora del dia per viatjar. Cadascun d'aquests canvis en la demanda provoca canvis al seu torn en els corresponents punts d'equilibri. En termes de modelització, el nou patró de flux produeix nivells de servei per a rutes i modes que poden o no ser coherents amb els assumits a l'hora d'estimar la matriu de viatge fix. Això requereix una re-estimació de la matriu i, per tant, alimentar els nous nivells de servei en el procés d'estimació per obtenir-ne una de nova. És possible que el procés es repeteixi sistemàticament fins que s'obtinguin les matrius de desplaçament amb els valors dels costos de viatge que siguin compatibles amb els fluxos estimats per a cada xarxa. Aquest nivell superior s'anomena l'equilibri del sistema enfront de l'equilibri de la xarxa.

Definicions i notacions

S'ha d'introduir una notació més segons es requereixi, però els elements bàsics utilitzats en aquesta etapa són:

- T_{ijr} és el nombre de viatges entre i i j per la ruta r ,
- V_a és el flux de l'arc a en vehicles per hora (VPH) o unitats de passatgers per cotxe (pcu) per hora, on normalment un autobús equival a entre 2 i 3 pcu i camions entre 3 i 4 pcu,
- $C(V_a)$ és la relació cost-flux de l'arc a ,
- $c(V_a)$ és cost real d'un determinat nivell de flux V_a ; el cost quan $V_a = 0$ es coneix com a cost de flux lliure,
- c_{ijr} és el cost de viatjar des de i a j per la ruta r ,

$$\delta_{ijr}^a = \begin{cases} 1 & \text{si l'enllaç està en el camí o ruta } r \text{ de } i \text{ a } j \\ 0 & \text{en cas contrari} \end{cases}$$

S'utilitza un superíndex n per indicar una iteració particular en mètodes iteratius. S'utilitza un superíndex $*$ per indicar un valor òptim.

En molts casos, és important reconèixer que hi ha diferents usuaris de la carretera i que poden mostrar comportaments diferents en el mateix arc. Per tant, es sol introduir un índex addicional (normalment u) per a la classe d'usuari. És possible tenir diferents classes d'usuari per a cada tipus de vehicle (cotxe, autobús, camió) i per a diferents tipus de

conductors en funció del seu motiu de viatge, disposició a pagar (ingressos) peatges i arquejar, i altres característiques personals rellevants per a estudiar.

Mètodes d'assignació trànsit

Els principals objectius de l'etapa clàssica d'assignació de trànsit són:

1. Principal:

- per obtenir bones mesures de xarxa agregades, p. ex. flux total d'autopistes, ingressos totals en servei d'autobús;
- estimar els costos de viatge zona a zona (temps) per a un nivell de demanda determinat;
- per obtenir fluxos d'arcs raonables i per identificar arcs fortament congestionats.

2. Secundari:

- estimar les rutes utilitzades entre cada parella O-D;
- per analitzar quines parelles O-D fan servir un arc o una ruta particular;
- per obtenir moviments de gir per al disseny de futures unions.

En termes generals, s'assoleixen els objectius principals amb més precisió que els secundaris. Això és bàsicament perquè aquests models tenen més probabilitats d'estimar un valor agregat que desagregar els valors correctament.

Les entrades bàsiques requerides per als models d'assignació són:

- Una matriu de viatges que expressa la demanda estimada. Normalment, aquesta és una matriu d'hora punta en àrees congestionades urbanes. Les mateixes matrius poden estar disponibles en termes de viatges de persones; per tant, s'han de convertir en viatges de vehicles en la mesura que es descriuen les relacions de capacitat i flux de velocitat en aquests termes.
- Una xarxa, és a dir, arcs i les seves propietats, incloses les corbes de velocitat de circulació.
- Principis o regles de selecció de rutes que es consideren rellevants per al problema en qüestió.

Els mètodes d'assignació de trànsit impliquen un conjunt de regles sobre com identificar rutes desitjables per connectar l'origen a la destinació i després una forma sistemàtica d'assignar viatges OD a aquestes rutes per aconseguir certes característiques de la realitat.

Elecció de ruta

La premissa bàsica en l'assignació és l'assumpció d'un viatger racional, és a dir, triar una ruta que ofereix els costos individuals menys percebuts (i anticipats). Es pensa que hi ha diversos factors que influeixen en l'elecció de la ruta quan es condueix entre dos punts; inclouen el temps de viatge, la distància, el cost monetari (combustible i altres), la congestió i les cues, el tipus de maniobres requerits, el tipus de carretera (autopista,

carretera troncal, carretera secundària), paisatges, senyalització, obres viàries, fiabilitat del temps de viatge i hàbit. La producció d'una expressió de costos generalitzada que incorpori tots aquests elements és una tasca difícil. A més, no és pràctic intentar modelar tots ells en un model d'assignació de trànsit i, per tant, les aproximacions són inevitables.

L'aproximació més habitual és considerar només dos factors en l'elecció de ruta: temps i cost monetari. A més, sovint es considera que el cost monetari és proporcional a la distància de viatge. La majoria dels programes d'assignació de trànsit permeten a l'usuari assignar pesos per viatjar amb temps i distància per tal de representar la percepció dels conductors d'aquests dos factors. La suma ponderada d'aquests dos valors es converteix en un cost generalitzat utilitzat per estimar l'elecció de la ruta. Hi ha proves que suggereixen que, almenys pel trànsit urbà d'automòbils, el temps és el factor dominant en l'elecció de la ruta. No obstant això, fins i tot si es combina el temps i distància en una funció de cost generalitzada, només es pot explicar quelcom de l'ordre del 60 al 80% de les rutes realment observades a la pràctica. Com que l'aportació marginal d'altres factors en l'elecció de ruta desenfrenada és molt reduïda, la part inexplicable s'ha d'atribuir a factors com les diferències de percepció, la informació imperfecta sobre els costos de la ruta o simplement errors.

El fet que els diferents conductors sovint triïn diferents rutes quan es viatja entre els dos punts es pot atribuir a tres tipus diferents de motius:

1. Diferències en les percepcions individuals del que constitueix la 'millor ruta'; potser alguns vulguin reduir al mínim temps, altres consums de combustible i moltes combinacions d'ambdós, i això introdueix una varietat en les opcions de ruta.
2. El nivell de coneixement de les rutes alternatives varia i això presenta una aparent irracionalitat (des del punt de vista de l'observador) en les eleccions.
3. Els efectes de la congestió afecten primer les vies més curtes i fan que els seus costos generalitzats siguin comparables a les rutes inicialment menys atractives.

Normalment, es fa servir el primer motiu a través de diverses classes d'usuaris, el segon a través 'd'efectes estocàstics' i el tercer a través de l'assignació i l'equilibri congestionats.

Assignació de transport públic

En aquesta secció es resumeix la problemàtica associada a l'elecció i assignació de ruta per als passatgers que utilitzen xarxes de transport públic. Aquests problemes són, en molts aspectes, més difícils que els trobats per l'assignació de transport privat; els requisits informàtics tendeixen a ser més pesats i fins i tot els millors mètodes requereixen importants suposicions simplificadores. Els últims anys han vist millores significatives en les tècniques d'assignació de trànsit que condueixen a una millor prestació de serveis de transport públic i eficiència operativa.

Primer es discuteixen les qüestions que fan que l'assignació de transport públic sigui diferent de l'elecció de la ruta del vehicle privat. A continuació, es detallen alguns dels problemes en l'assignació de transport públic.

Oferta

La xarxa de serveis de transport públic és diferent de la dels cotxes particulars. Inclou, com a arcs, seccions dels serveis d'autobús o ferrocarril que s'executen entre dues parades o estacions. El concepte de capacitat de connexió ara està relacionat amb la capacitat de cada unitat (autobús, tren) i la seva freqüència corresponent. El temps de viatge té un component en el vehicle, així com components per esperar en parades i caminar des de, i cap a elles. Moltes de les seccions de transport públic utilitzen arcs viaris.

Passatgers

En l'elecció de la ruta de transport públic es tracta del moviment de passatgers i no de vehicles. Els passatgers poden caminar cap a una parada, intercanviar entre dos serveis i, fins i tot, conduir part del camí per abordar un servei de transport públic més tard. Això requereix la necessitat de proporcionar i especificar caminar i transferir arcs entre diferents serveis, diferents modes de transport públic (autobús, ferrocarril) i entre instal·lacions de transport públic i privat.

Costos monetaris

A les xarxes privades d'automòbil generalment es suposa que el cost monetari està directament associat al consum de combustible, que al seu torn és directament proporcional a la distància de viatge. Són ambdues aproximacions, però s'accepten generalment, ja que els conductors no perceben aquests costos de forma tan directa com el passatger que compra un bitllet quan comença un viatge en autobús. Els sistemes de pagament moderns basats en targetes intel·ligents o telèfons mòbils permeten estructures de tarifes més complexes i s'han introduït en moltes operacions de transport públic: tarifes variables amb distància, tarifes planes (independentment de la distància recorreguda), tarifes horàries (per a una o més zones geogràfiques específiques), bitllets de combinació i transferència (vàlids per a dos o més serveis), tarifes de temps (per exemple, vàlides per a qualsevol nombre d'embarcacions en una hora), diaris, setmanals i altres entrades de temporada per a un servei fix o que cobreix una o més zones i modes. Aquesta àmplia gamma de tarifes estableix uns requisits difícils sobre l'elecció de la ruta i els models d'assignació, ja que els costos monetaris no depenen directament de la distància sinó, en general, de la ubicació de l'origen i la destinació, i de la ruta elegida.

2. Mètodes de recollida de dades de transport

2.1. Mètodes tradicionals pels patrons de mobilitat

Els mètodes tradicionals (Ortuzar and Willumsen 2011) són els relacionats amb les enquestes de transport. Tanmateix, les limitacions pràctiques tenen una gran influència en la determinació del tipus d'enquesta més adequat per a una situació i àmbit concret. L'especificació del model desitjat i el disseny d'un pla d'enquestes no són una qüestió senzilla i requereixen coneixements i experiència considerables.

Els tipus d'enquestes més utilitzades són:

- Enquesta Domiciliària familiar: els viatges realitzats per tots els membres de la llar per tots els modes de transport, tant dins de l'àrea d'estudi com de sortida/arribada a la zona durant el període de l'enquesta. Aquesta enquesta ha d'incloure informació socioeconòmica (ingressos, propietat del cotxe, mida i estructura de la família, etc.). Aquesta informació és molt eficient a l'hora de generar dades que permeten estimar models de generació de viatges i models de modes; a més, les dades sobre viatges a la llar proporcionen una bona informació sobre la distribució de les longituds de viatge a la ciutat, un element important en l'estimació dels models de distribució de viatges.
- Enquesta Domiciliària Individual: Selecció d'individus concrets a partir d'una descripció del marc de mostreig basat en els domicilis.
- Enquestes d'intercepció o cordó extern: Recullen dades sobre persones que travessen la frontera de l'àrea d'estudi, en particular els no residents de l'àrea d'estudi. Aquestes dades també es poden utilitzar per comprovar i amplificar les dades d'enquestes domiciliàries en punts de pas concrets de l'àrea d'estudi, ja que sol haver una petita quantitat de dades recollides, fins i tot en una enquesta molt gran. Es tracta d'enquestes més breus, realitzades en punts que intercepten els viatges que arriben i surten de l'àrea d'estudi: enquestes fora de límit, a bord de vehicles de transport públic o en punts d'intercanvi de manera (és a dir, aeroports).
- Enquestes d'intercepció, cordons interns i línies de pantalla: es necessiten per mesurar els viatges per part de no residents, i de nou per verificar les dades de la font domiciliària, fins a cert punt. Són entrades importants per a altres models.

- Recompte de persones i vehicles: són de baix cost i es requereixen per al calibratge, la validació i altres comprovacions a altres enquestes.
- Enquestes de temps de viatge: es requereixen per calibrar i validar la majoria de models i poden ser necessàries tant per a viatges de transport públic com per privat.
- Altres dades relacionades: per crear models de predicció robustos per grans àrees metropolitanes, és important disposar d'una metodologia d'enquesta que permeti la integració d'elements de dades relacionats que influeixen en el comportament del viatge. S'inclou:
 - Inventari d'ús del sòl; zones residencials (densitat d'habitatge), zones comercials i industrials (per tipus d'establiment), places d'estacionament, etc. Aquests són especialment útils per als models de generació de viatges.
 - Infraestructures i inventaris de serveis existents (xarxes de transport públics i privats, tarifes, freqüència, etc.; ubicació i temps de senyal de trànsit); Aquests són essencials per al calibratge del model, especialment els models de distribució i assignació.
 - Informació d'enquestes especials sobre actituds i elasticitat de la demanda, com les enquestes de Preferències Declarades.
- Enquestes de Preferències Declarades: són un conjunt de metodologies que es basen en dades declarades pels individus sobre com actuarien davant de diferents situacions hipotètiques que li són presentades i que han de ser les més realistes possible.

2.2. Recollida de dades de sensors

La recollida precisa de dades de trànsit permet als planificadors implementar sistemes intel·ligents de transport, incorporant els sensors de trànsit en algunes vies de la xarxa viària. Encara que hi ha diverses opcions de recollida de dades convencionals en ruta, ofereixen una cobertura limitada i tenen altres inconvenients i limitacions. Per gestionar l'augment constant del trànsit vehicular, les autoritats i els desenvolupadors necessiten un sistema de gestió del trànsit precís. Hi ha dos grups de recollida de dades de sensors de trànsit:

1. Recollida de dades tradicionals: sensors de trànsit a les vies, hi ha els següents tipus:
 - Tub pneumàtic de carretera: són sensors de tub de cautxú buits per carretera que recullen dades de trànsit i velocitat. En aquest procés, es col·loquen diversos tubs al llarg de la carretera amb un dels seus extrems units a un comptador/classificador de trànsit. Quan un vehicle passa sobre el tub, exerceix pressió sobre ell, provocant l'inductor en el classificador de trànsit, registrant així el recompte. En general, cinc sistemes pneumàtics de tubs de carretera s'utilitzen en enquestes de trànsit a curt termini que duren una setmana o dues. Encara que els tubs són relativament econòmics que altres sensors de

carretera, es poden deteriorar fàcilment, la qual cosa pot afectar la precisió de les dades que es recullen.

- Sensors piezoelèctrics: són una de les opcions més precises disponibles per mesurar el cop d'eix i normalment s'utilitzen per calcular els recomptes dels vehicles, la classificació del trànsit i el seguiment del compliment de la velocitat.
 - Bucles d'inducció: són una de les tecnologies de recollida de dades de trànsit més habituals, que s'utilitzen en el sistema de transport intel·ligent. El bucle és un corrent continu de cables aïllats connectats a un detector de vehicles. Per a la recopilació de dades de trànsit, diversos bucles indicatius són incrustats en vials a través dels quals es passa un corrent elèctric, que crea un camp electromagnètic. El detector registra el recompte, ja que el vehicle passa pel camp electromagnètic.
2. Recollida de dades progressives: sensors de trànsit al costat de la via, hi ha els següents tipus:
- Sensors de tràfic acústic: recullen dades de trànsit mitjançant la identificació del soroll dels pneumàtics. També mesura la velocitat del vehicle per les vies de diversos carrils, fins i tot en condicions d'alta circulació. Segons diversos estudis independents, els sensors acústics proporcionen un recompte de trànsit precís, fins i tot si les velocitats dels vehicles varien amb un marge notable. A més, els sensors acústics de trànsit representen una alternativa econòmica als sensors convencionals en carretera, especialment en aquells espais on les condicions de circulació i de trànsit limiten la vida útil del sensor de carretera.
 - Radar de microones: es coneix com a radar Doppler que utilitza ones de freqüència de ràdio per mesurar el flux del trànsit, inclosa la classe del vehicle (efecte Doppler). L'avantatge d'utilitzar radar de microones és que es pot operar a una longitud d'ona d'alta freqüència, independentment de les condicions climàtiques. Mentre que el desavantatge és l'oclusió del vehicle (es pot perdre un vehicle a causa d'un vehicle més gran).
 - Detecció de la imatge de vídeo (VID): és un altre notable desenvolupament en el camp de la recollida de dades de trànsit. És un sistema avançat de vigilància del trànsit que processa entrades d'una càmera de vídeo per controlar el trànsit en temps real. Tot i que és un dels mètodes més recents per a la recopilació de dades de trànsit, la característica de classificació del vehicle està limitada a les hores de llum, tret que el lloc de gravació estigui ben il·luminat. Són sensors cars d'implantació, manteniment i explotació

2.3. Mètodes actuals

La recollida de dades de mobilitat i transport en l'actualitat (Chen et al. 2016) està relacionada amb les noves fonts de tecnologies intel·ligents d'informació i comunicació

(TIC), que generen gran conjunt de dades. Alguns exemples d'aquestes noves fonts de dades són:

- Telefonia mòbil
- Seguiment de flotes (taxis, distribuïdors de mercaderies, etc.)
- Targetes intel·ligents de transport
- Xarxes socials (Twitter, Facebook, etc.)

Aquest treball està enfocat a l'anàlisi de les matrius OD de mobilitat, estimades a partir de dades de mobilitat procedents de la telefonia mòbil generades passivament. En l'anàlisi dels patrons de mobilitat humana, les dades de telefonia mòbil generades passivament són emprades molt freqüentment en l'actualitat (des de fa uns 5-7 anys), ja que altres fonts esmentades no poden capturar l'espectre complet del patró de mobilitat d'un individu durant diversos dies, que impliquen l'ús de múltiples modes de transport i la realització de desplaçament per diferents motius.

C. Chen et al./Transportation Research Part C 68 (2016) 285–299

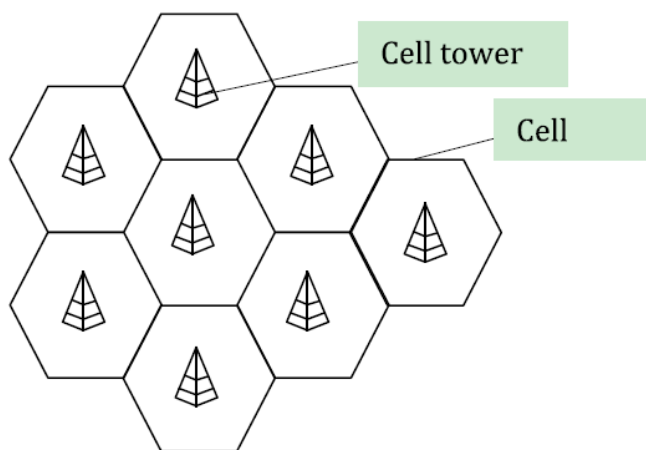


Figura 2.1: Exemple de la xarxa cel·lular

2.3.1. Dades telefonia mòbil

La informació obtinguda de les dades passives de mòbils (Chen et al. 2016) es genera com a resultat de la comunicació d'un telèfon amb la xarxa mòbil mantinguda i operada per operadors de xarxes mòbils; aquest procés s'anomena posicionament. Una xarxa cel·lular (Figura 2.1) és la que permet que els telèfons mòbils es comuniquen entre si; comprèn diverses estacions base o torres, cadascuna servint una àrea diferent, que es denomina cel·la. Cada cel·la té un identificador de cel·la únic.

El posicionament del telèfon mòbil es requereix quan el dispositiu d'un usuari es comunica amb la xarxa. Quan un usuari inicia un esdeveniment de connexió a la xarxa, l'operador de la xarxa mòbil ha de conèixer la seva ubicació per tal de determinar la torre de les cel·les que s'utilitza per canalitzar aquest esdeveniment. D'aquesta manera, es generen les dades de posicionament quan es produeix un esdeveniment. Aquestes dades

es generen automàticament i passivament per a propòsits propis dels operadors de xarxes mòbils, incloent-hi la recollida d'informació de facturació i la gestió de la xarxa.

El posicionament pot donar lloc a dos tipus de dades de telefonia mòbil. Les dades de registres de detall de trucades (CDR) probablement són les més utilitzades avui en dia. Tots els registres d'una informació CDR representen una trucada telefònica, amb informació sobre l'autor de la trucada, l'hora de sortida de la trucada, la durada de la trucada i les coordenades XY de la torre que van canalitzar la trucada per primera vegada quan es va iniciar la trucada.

Un conjunt de dades de telefonia mòbil que s'utilitza amb menys freqüència s'anomena "dades d'observació". Es genera una observació cada vegada que es posiciona un telèfon. Es pot pensar que les dades d'observació són una versió processada de les dades CDR, però es diferencia de les dades CDR en dos aspectes:

- Resolució temporal: les dades d'observació poden tenir un major nivell de resolució temporal que les dades CDR, o sigui, les dades CDR solament generen un registre per trucada, en canvi les altres poden generar múltiples observacions.
- Resolució espacial: les dades d'observació poden tenir un major nivell de resolució espacial que les dades CDR. Les localitzacions descrites en CDR són ubicacions de la torre de telefonia cel·lular i, per tant, depenen de la densitat de la xarxa cel·lular. D'altra banda, les ubicacions que es registren en les dades d'observació són els resultats de la triangulació de múltiples torres i aquestes ubicacions són àmpliament vistes com a ubicacions de dispositius enfront de les ubicacions de la torre en les dades del CDR.

2.3.2. Dades de seguiment de flotes de taxis

Les trajectòries de taxis (Zhang et al. 2016) estan constituïdes per una sèrie d'ubicacions que poden reflectir quan i on es produeixen les activitats i, per tant, s'utilitzen per analitzar la dinàmica humana. Així, una major comprensió de la dinàmica humana ha motivat molts estudis sobre la gestió del transport, la planificació urbana, la modelització de l'ús del sòl i la predicció del flux de trànsit. Aquestes dades es recullen mitjançant una flota de taxis específica equipats amb dispositius GPS (Sistema de Posicionament Global). S'han proposat mètodes, com ara passeig aleatori en el cas de flotes de vehicles flotants adreçats al monitoratge en temps real de l'estat del trànsit, direcció aleatòria, etc.

Per expressar la distribució espacial dels patrons de mobilitat, es divideix la xarxa viària en punts d'encreuament, generant petits trossos de camí, anomenats segments. Cada segment rep un número d'identificació únic. La majoria d'estudis inicials van investigar la distribució espacial de trajectòries basades en regions i noms de carreteres. Menys estudis s'han centrat en l'escala més fina dels segments en una xarxa viària urbana.

Un dels estudis realitzats amb aquest tipus de dades és a Wuhan, que és la ciutat més gran de la Xina central. L'estudi es troba descrit a: (Zhang et al. 2016).

La distribució espacial mostra diferents destinacions de viatge a través de tècniques de modelatge temàtic, com ara l'assignació Latent Dirichlet (LDA) en nivell de segment. Es van extreure recorreguts amb diferents trajectòries a partir d'un algoritme d'extracció dels viatges a partir de dades GPS crues. Tanmateix, la ubicació de les trajectòries de taxi no pot expressar explícitament patrons de mobilitat humana, doncs en general no representen una mostra representativa del patró general de mobilitat. És pre-processen les dades de la trajectòria crua dels taxis mitjançant l'addició d'informació semàntica abans d'utilitzar LDA per donar a les ubicacions identificadors únics que es poden entendre. D'aquesta manera, un número d'identificació de segments s'ha associat a ubicacions geogràfiques corresponents mitjançant un procés de correspondència de mapes (map matching). Les trajectòries de taxi es transformen en una sèrie de trajectòries semàntiques que consistien en destinacions de viatge amb nombres d'identificació de segments. Cada trajectòria semàntica de taxis és anàloga a un document amb una sèrie d'ID de segment, mentre que les trajectòries semàntiques de taxis massius s'agrupen com un corpus per identificadors de segments. Els segments amb identificadors de segment similars s'han d'agregar per temes que expressen un significat ocult comú, utilitzant el mètode LDA. Posteriorment, es pot utilitzar aquests temes per analitzar la similitud espacial dels patrons de mobilitat i la seva relació amb les activitats diàries o motius dels viatges.

2.3.3. Dades de targetes intel·ligents de trànsit

La targeta intel·ligent (Long and Thill 2015) que registra la informació completa del viatge de la targeta convencional és una forma alternativa de tecnologia d'adquisició d'ubicació. Els sistemes de recollida de tarifes automatitzades de targetes intel·ligents s'apliquen cada vegada més en sistemes de trànsit públic. Simultàniament amb la recaptació d'ingressos, aquests sistemes poden produir patrons de viatges intensius de titulars de targetes, dades que són útils per analitzar dinàmiques urbanes. L'ús de targetes intel·ligents s'ha convertit en significatiu a causa del desenvolupament d'Internet i l'augment de la complexitat de les tecnologies de la comunicació mòbil. Les dades generades pels sistemes de targetes intel·ligents fan un seguiment de les transaccions a bord detallades de cada titular de la targeta. S'argumenta que la tecnologia de la targeta intel·ligent pot proporcionar informació valuosa perquè és una tècnica contínua de recollida de dades que proporciona un diari de viatge complet i en temps real per a tots els viatgers d'autobús. Les dades de la targeta intel·ligent (SCD) es poden utilitzar per validar els models tradicionals de viatge aplicats al transport públic; aquestes targetes són molt utilitzades a la Xina (per exemple a Beijing) o Hong Kong. En contrast amb la recopilació de dades de targetes intel·ligents, les enquestes de viatges convencionals tenen els inconvenients de ser cares i poc freqüents. Notablement, el trànsit SCD recopila dades fonamentalment de la mateixa manera que el sistema AVI (identificació automàtica de vehicles), que ha estat àmpliament utilitzat als EUA per identificar automàticament els vehicles. AVI s'utilitza en alguns estats dels EUA amb finalitats de planificació. Els resultats de la identificació es validen mitjançant dades d'enquestes domiciliàries. S'afirma que la SCD és un bon substitut de les enquestes de domiciliàries o almenys un complement.

2.3.4. Dades de xarxes socials

Els serveis d'intercanvi d'ubicacions (LSS) (Cheng et al. 2011) com Foursquare, Gowalla i Facebook proporcionen centenars de milions de petjades controlades per l'usuari (p. ex., "Checkins"). Aquestes petjades a escala global proporcionen una oportunitat única per estudiar les característiques socials i temporals de la manera com les persones utilitzen aquests serveis i modelar l'activitat humana que són factors importants per al disseny de futurs serveis de telefonia mòbil + localització, previsió de trànsit, planificació urbana, ubicació de nous comerços, així com models epidemiològics de propagació de malalties.

A diferència de les dades dels telèfons mòbils i les trajectòries derivades dels rastrejadors de GPS, les comprovacions tenen diverses característiques úniques:

- Són inherentment socials, ja que els usuaris revelen la seva ubicació als seus amics, el que significa que es pot observar directament l'estructura social i el seu impacte en la mobilitat humana.
- Les comprovacions estan associades a llocs concrets (p. ex., un restaurant), que permeten una major anàlisi del tipus de lloc.
- Els controls es poden augmentar amb missatges curts, proporcionant una visió parcial dels pensaments i motivacions dels usuaris d'aquests serveis.

Alguns exemples són:

- Els investigadors de Facebook van analitzar la 'distància' entre les relacions socials dels usuaris de Facebook i van utilitzar ubicacions dels amics d'un usuari per predir la ubicació geogràfica de l'usuari (Backstrom, Sun, and Marlow 2010).
- (Cheng, Caverlee, and Lee 2010) va modelar la distribució espacial de paraules en el contingut generat per l'usuari de Twitter per predir la ubicació de l'usuari.
- Es pot estudiar la caracterització de les propietats de la xarxa de relació amb la geografia local (Yardi and Boyd 2010).
- S'ha estudiat el comportament de l'usuari pel que fa al camp d'ubicació dels perfils d'usuari de Twitter (Hecht et al 2011).
- (Lindqvist et al. 2011) van analitzar com i per què la gent utilitza serveis d'intercanvi d'ubicacions i va discutir sobre els problemes de privadesa relacionats amb els serveis d'intercanvi de llocs.
- A més de localitzacions, els investigadors també han explorat dinàmiques temporals relacionades amb activitats socials en línia (Long and Thill 2015)(Long and Thill 2015)(Long and Thill 2015) (Golder, Wilkinson, and Huberman 2007).

3. Dades de demanda de Kineo

3.1. Introducció

Kineo és una empresa tecnològica que analitza i visualitza dades de geolocalització per a la presa de decisions en diferents sectors: transport, turisme, geomarketing, estudis ambientals, etc. Aquesta empresa utilitza les TIC (Tecnologies de la Informació i les Comunicacions) per la recollida automàtica de grans quantitats de dades espai-temporals que obren noves oportunitats per generar estalvis de costos i temps per a autoritats públiques i empreses privades que participen en la planificació i la gestió del transport.

Kineo ofereix una solució que extrau i analitza les dades de geolocalització anònimes dels dispositius de telefonia mòbil (companyia Orange) i la combina amb altres fonts de dades (dades socioeconòmiques del padró i en el futur dades de Twitter, SCD, càmeres de vigilància, etc.) per proporcionar informació exhaustiva i actualitzada sobre viatges per motius en els diferents intervals del temps i amb un cost molt menor, tant en temps, com en diners, del que requereixen els mètodes tradicionals basats en enquestes.

Aquesta empresa segueix unes pautes genèriques de la mineria de dades a l'hora d'analitzar les dades: pre-processament i emmagatzematge, neteja de dades, extraccions de patrons d'activitat i mobilitat, i modelització estadística per a l'expansió de la mostra. Un cop fet això, les dades les transformen en matrius de mobilitat OD (origen-destinació) segons l'especificació de la Zonificació requerida pel client o en una distribució dinàmica de població i estadístiques d'activitat.

3.2. Dades

L'origen de les dades que constitueix l'entrada principal de la solució de Kineo són registres de telefonia mòbil anònims, que s'obtenen mitjançant acords de col·laboració amb operadors de xarxes mòbils, malgrat que els acords permanents els tenen només amb Orange per Espanya i amb Entel per Xile. Segons les dades disponibles de cada operador, la solució pot funcionar amb dos tipus de registres de dades:

- Registres de detall de trucades (CDR), emmagatzemats per operadors mòbils amb finalitats de facturació.

- Dades de la sonda de xarxa, generades per a la supervisió i la gestió de xarxes.

Els CDR es produeixen cada vegada que un telèfon mòbil interactua amb la xarxa mitjançant una trucada de veu, un missatge de text o una connexió de dades d'Internet. Entre altres informacions, cada CDR conté un identificador de l'usuari, una marca de temps i una ubicació. Gràcies a la creixent utilització de telèfons intel·ligents i serveis de dades, la majoria dels usuaris generalment generen un registre com a mínim cada 20 o 30 minuts sense interacció amb el telèfon mòbil. En els últims anys, alguns operadors han començat a emmagatzemar les dades generades per la sonda de xarxa. La sonda de xarxa no només registra la ubicació de l'usuari per als esdeveniments actius inclosos en els CDR, sinó que també genera registres periòdics sense necessitat d'una transacció de telecomunicacions, proporcionant així una major resolució temporal. Tant en CDR, com en dades de sonda de xarxa, la posició sol donar-se per la cel·la de la xarxa on es troba l'usuari, que proporciona una resolució relativament alta en àrees urbanes denses, però una ubicació menys precisa a les zones rurals. Recentment, alguns operadors estan començant a millorar la resolució espacial, emmagatzemant les coordenades geogràfiques dels usuaris estimades, mitjançant triangulació. Les dades proporcionades per l'operador de telefonia també poden incloure informació sociodemogràfica vinculada als usuaris, com l'edat i el sexe.

A més de les dades de telefonia mòbil, Kineo ha incorporat el contrast amb dades sociodemogràfiques externes (p. ex., dades del cens o registres administratius), dades d'ús del sòl i dades a la xarxa de transport (rutes, temps de viatge, horaris, etc.) per fer l'expansió dels perfils de mobilitat dels clients de la companyia de telefonia a tota la població.

L'extracció de l'activitat i la informació de la mobilitat dels registres de telèfons mòbils es basa en un conjunt d'algoritmes patentats i desenvolupats per Kineo. El procés consisteix en els següents subprocessos:

1. Pre-processament i neteja de dades
2. Generació de la mostra efectiva potencial
3. Identificació del lloc de residència
4. Selecció de la mostra efectiva
5. Extracció d'activitats i viatges diaris
6. Expansió de la mostra a la població total
7. Generació d'informació de sortida

3.2.1. Matriu de dades

Les matrius de dades proporcionades per Kineo per a realitzar aquest projecte estan dividides en submatrius segons el lloc de residència i el motiu de viatge. Les dades subministrades per la caracterització de la mobilitat de la Primera Corona Metropolitana consisteixen en 3 grups de matrius segons el lloc de residència (Barcelona, Primera Corona i Regió Metropolitana de Barcelona), i en cada grup, es va subministrar una submatriu per a cada motiu de viatge (casa, feina, freqüent i poc freqüent). Cadascuna

d'aquestes es representen en arxius amb 9,3 milions de registres, per tant sí per a cada lloc de residència hi ha 4 motius, això suma un total de 111,6 milions registres complexos. Cada registre té una caracterització de nombre total de viatges (tots els modes) per cada hora del dia d'un dia feiner mig de l'any 2017 i segment demogràfic (caracteritzat per grup d'edat i gènere). Es van emprar dades de telefonia corresponent als tres darrers dies del mes de Març del 2017. La població d'estudi són tots els individus de la Regió Metropolitana de Barcelona que realitzen algun desplaçament relacionat amb la Primera Corona Metropolitana.

El format dels registres que caracteritzen les matrius de mobilitat és el següent:

- Origen: zona d'origen del viatge
- Destinació: zona de destinació del viatge
- Temps: franja horària en què s'inicia el viatge
- Total: total de viatges d'origen a destinació iniciats en temps
- Home_16-24: total de viatges d'origen a destinació iniciats en temps i realitzats per homes amb edat compresa entre 16 i 24 anys.
- Dona_16-24: total de viatges d'origen a destinació iniciats en temps i realitzats per dones amb edat compresa entre 16 i 24 anys.
- Home_25-44: total de viatges d'origen a destinació iniciats en temps i realitzats per homes amb edat compresa entre 25 i 44 anys.
- Dona_25-44: total de viatges d'origen a destinació iniciats en temps i realitzats per dones amb edat compresa entre 25 i 44 anys.
- Home_45-64: total de viatges d'origen a destinació iniciats en temps i realitzats per homes amb edat compresa entre 45 i 64 anys.
- Dona_45-64: total de viatges d'origen a destinació iniciats en temps i realitzats per dones amb edat compresa entre 45 i 64 anys.
- Home_65-99: total de viatges d'origen a destinació iniciats en temps i realitzats per homes amb edat compresa entre 65 i 99 anys.
- Dona_65-99: total de viatges d'origen a destinació iniciats en temps i realitzats per dones amb edat compresa entre 65 i 99 anys.

En aquest treball la caracterització que fa referència a edat i sexe no s'ha considerat per fer l'anàlisi per tal de limitar el volum de dades a tractar conjuntament, amb les matrius desagregades per lloc de residència i motiu de viatge, i les variables d'origen, de destinació, de viatges totals, i per franges horàries ja es tenia un gran volum de dades per realitzar l'estudi de validació de la qualitat dels patrons de mobilitat estimats a partir de dades de la telefonia mòbil.

4. Projecte Virtual Mobility LAB

L'objectiu del projecte Virtual Mobility LAB (Barceló, Montero, & Ros-Roca, 2018) és proporcionar un marc comú per donar suport en la presa de decisions sobre el disseny i l'avaluació dels impactes dels nous conceptes de transport i tipus de vehicles que sorgeixen com a conseqüència de la combinació de la dinàmica urbana, els canvis en el comportament dels viatgers i les aplicacions de les TIC (Tecnologies d'Informació i Comunicació). Aquest entorn de modelització de la mobilitat i el sistema de transport de la Primera Corona Metropolitana, està adreçat a avaluar els nous conceptes i formes de mobilitat futura, que es classifiquen en dues categories:

- Conceptes de mobilitat basats en les properes tecnologies de l'automoció:
 - Vehicles elèctrics
 - Vehicles connectats
 - Vehicles autònoms (sense conductor)
- Conceptes de mobilitat basats en nous sistemes de transport recolzats per aplicacions TIC:
 - Lloguer de cotxes per hores (*car-sharing*), compartir cotxe (*ride-sharing*), transport a la de passatgers (*multiple passenger ride-sharing*) i altres formes de mobilitat com a servei (MaaS, Mobility as a Service) que s'estan proposant en l'actualitat, molt relacionades amb el concepte de transport a la demanda.
 - Noves formes de multi-modalitat, realització d'un desplaçament en diverses etapes on cadascuna d'elles usa un mode específic de transport, convencional o innovador (patinets per la darrera milla, etc.).
 - Nodes d'intercanvi multimodals i gestió d'estacionaments.

El Virtual Mobility LAB és una plataforma integrada de simulació multimodal adreçada a la presa de decisions per part d'autoritats del transport i operadors en relació al disseny i avaluació d'impactes de nous conceptes de mobilitat, per a escenaris que evolucionen amb el temps.

4.1. Un enfocament metodològic per construir el nucli de simulació del “Virtual Mobility Lab”

El desenvolupament del model base actual (Barceló, Montero, and Ros-Roca 2018) s’ha efectuat sobre la plataforma comercial Visum (PTV). “Virtual Mobility Lab” és un cas d’ús de construcció de models *data-driven* per les fonts de dades emprades i per les eines i utilitats disponibles a Visum. El procés comença per identificar la regió geogràfica a modelar, que és l'àrea d'estudi, en aquest cas la Primera Corona Metropolitana (18 municipis). Els passos metodològics del procés es descriuen en el diagrama de flux de la Figura 4.1.

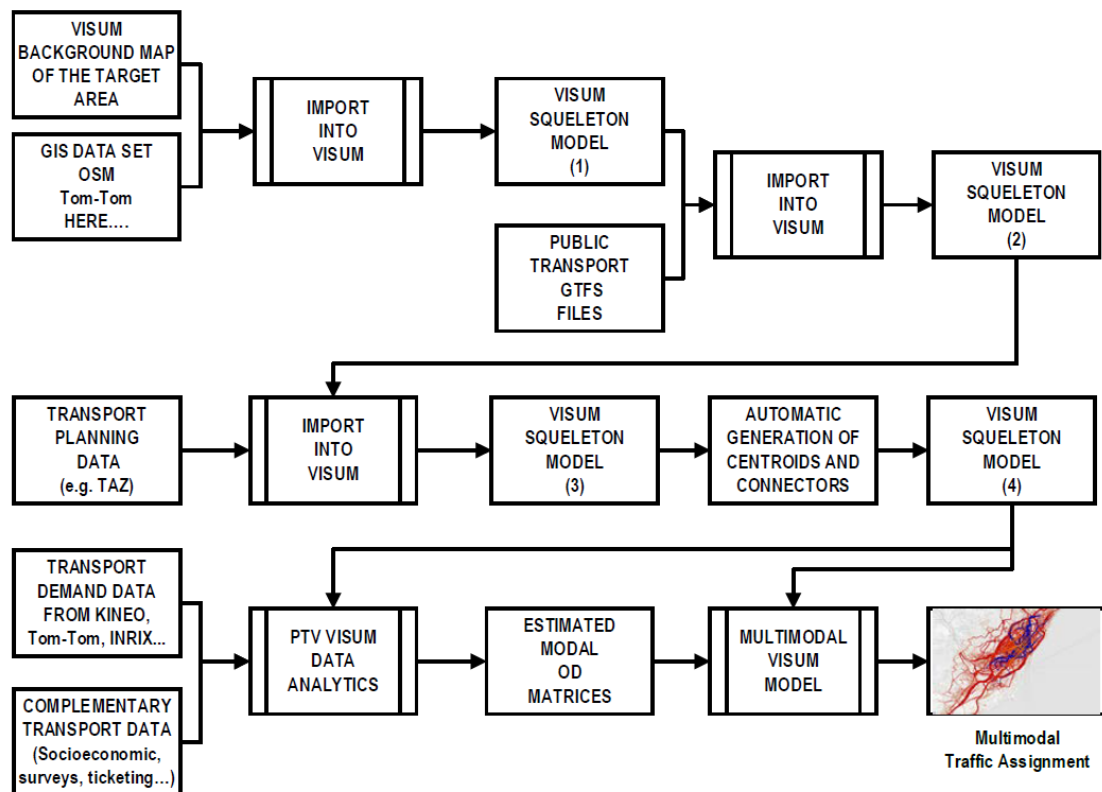


Figura 4.1: Diagrama del flux de treball del “Procés de construcció de models dirigits a dades”.

Font:(Barceló, Montero, and Ros-Roca 2018)

4.2. Model de la Primera Corona de l’Àrea Metropolitana de Barcelona

Seguint l’esquema anterior (Figura 4.1), el primer pas ha estat la selecció de l’àrea d’estudi. On la mostra Figura 4.2 les àrees inicialment candidates:

- Regió metropolitana de Barcelona (fronteres blaves)
- Àrea metropolitana de Barcelona (fronteres vermelles)
- Primera Corona de l’Àrea metropolitana de Barcelona (fronteres verdes)
- Ciutat de Barcelona (àrea interior de color gris fosc)

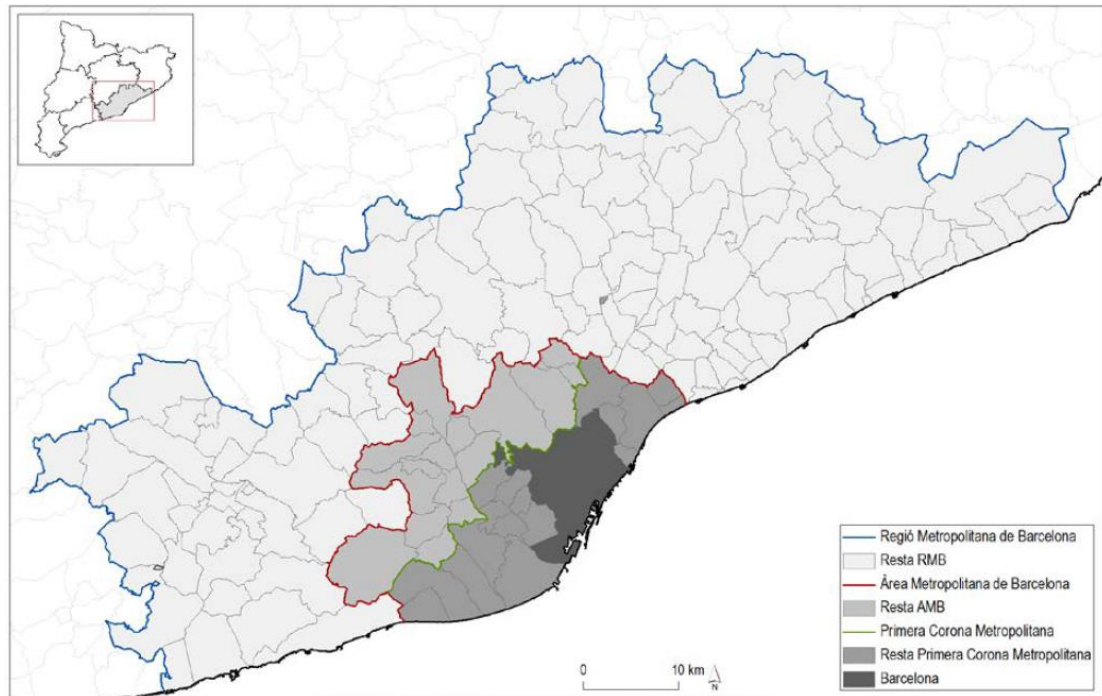


Figura 4.2: Candidats a l'àrea d'orientació del model Visum per al "Virtual Mobility LAB".

Font: (Barceló, Montero, and Ros-Roca 2018)

La decisió de l'àrea d'estudi per una primera fase d'implementació del Virtual Mobility LAB ha estat seleccionar la Primera Corona de l'Àrea Metropolitana, formada per 18 municipis: Barcelona, Badalona, Sant Adrià del Besòs, Santa Coloma de Gramenet, L'Hospitalet de Llobregat, Tiana, Montgat, Montcada i Reixac, Esplugues de Llobregat, Sant Just Desvern, Sant Feliu de Llobregat, Sant Joan Despí, Cornellà de Llobregat, El Prat de Llobregat, Sant Boi de Llobregat, Viladecans, Gavà i Castelldefels. Aquesta decisió no impedeix una futura expansió del model ni a l'Àrea Metropolitana ni a la Regió Metropolitana. Els motius principals d'aquesta elecció són:

- **Població.** La Regió Metropolitana de Barcelona té 164 municipis i 4.747.035 habitants. L'Àrea Metropolitana de Barcelona, amb 36 municipis, té 3.239.337 habitants. La Primera Corona de l'Àrea Metropolitana, amb els 18 municipis esmentats anteriorment, té 2.836.925 habitants, per tant representa la concentració més gran de població, al voltant de la ciutat de Barcelona.
- **Oferta del transport.** Concentra la xarxa viària principal i la major oferta de transport públic (més de 200 línies d'autobusos, amb més de 4.000 parades, 10 línies de metro, 15 línies de ferrocarril, 9 de ferrocarrils de la Generalitat i 6 de Renfe) i les dues línies de tramvia.

Una vegada seleccionada l'àrea d'estudi, el següent pas en el procés metodològic ha estat la definició de zones d'anàlisi del transport (TAZ, Transport Analysis Zones). Atès que un dels objectius, d'acord amb el propòsit d'utilitzar nous mètodes per generar matrius OD, a partir de l'explotació de les dades de telèfon mòbil de l'operador de telefonia d'Orange per produir matrius OD. Llavors cal definir un sistema de zonificació amb tres objectius:

1. Una TAZ ha d'estar ben equilibrada en termes de població, dividint l'àrea d'estudi en zones amb població similar i desviacions estàndard reduïdes (coeficient de variació inferior al 30%). Seguint les recomanacions pràctiques habituals de l'anàlisi de la demanda de transport, s'ha utilitzat les xifres de població de referència de mitjanes de població properes a 5.000 habitants per zona.
2. Al mateix temps, no s'ha d'oblidar que les TAZ són més grans que les seccions censals, i és el resultat de l'agrupació de zones censals pertanyents exclusivament a un districte municipal. Per tant, no poden haver-hi zones de transport que parcialment recobreixin 2 o més districtes municipals simultàniament i per extensió 2 o més municipis.
3. Finalment, també cal tenir en compte les relacions entre les seccions censals i les cel·les telefòniques, per tal de no dividir les cel·les entre dues o més seccions censals.

5. Metodologia

Per a la realització d'aquest treball, en primer lloc s'ha realitzat una recerca exhaustiva per entendre el context de modelització del sector de la Modelització i Planificació del Transport en àmbits metropolitans. Els conceptes el marc teòric i la seva aplicació al desenvolupament pràctic. Per això, s'han consultat llibres específics de modelització del transport, articles i documents relacionats. Seguidament s'ha analitzat el projecte Virtual Mobility LAB, per adquirir familiaritat amb la zonificació de l'àmbit d'estudi i les característiques que defineixen l'oferta de transport i sobretot la demanda de transport en forma de matrius de mobilitat modal per motius.

L'objectiu del treball és validar si les matrius de mobilitat obtingudes a partir del tractament de dades de telefonia mòbil poden emprar-se de manera satisfactòria per la modelització de la demanda en un context de model UTPS. Una vegada definit els principals objectius específics del treball en termes d'indicadors significatius a caracteritzar i contrastar amb fonts públiques de mobilitat actual (EMEF, 2015), s'ha analitzat amb R-Studio les dades subministrades per l'empresa Kineo.

L'anàlisi que es realitza és per validar i entendre els indicadors relacionats amb els patrons de mobilitat.

Abans de començar l'estudi es realitza un preprocessament de les dades per observar que està tot correcte.

La primera part dels resultats es busquen indicadors globals i relacions multivariants entre àmbits de mobilitat, llocs de residència, motius dels desplaçaments i modes de transport, on s'ha inspirat en el document de resultats de l'anàlisi de dades de l'Enquesta de Mobilitat En dia Feiner (EMEF) (Autoritat del Transport Metropolità 2015) per crear una proposta metodològica ampliada amb eines estadístiques d'anàlisi no emprades per defecte en el sector.

Per realitzar les relacions mencionades anteriorment, s'ha seguit un guió que és: anàlisi territorial general, segons el motiu de viatge i el lloc de residència, i s'ha estudiat per la Regió Metropolitana de Barcelona i després per la Primera Corona de Barcelona; seguidament s'estudia el mode de transport desagregant l'anàlisi territorial, però només amb la matriu OD per motiu del viatge, dins d'aquí s'ha volgut profunditzar en la Primera

Corona i el transport motoritzat, ja que el no motoritzat té molt pes dins de les dades però sol realitzar-se a peu i fa canviar la perspectiva. S'han calculat indicadors de volum del transport públic i privat mitjançant una ràtio. Finalment, s'han analitzat les franges horàries per veure les hores punta, on les dades s'han desagregat per motiu de viatge i lloc de residència a la RMB. Tots els resultats es comparen amb finalitats de validació amb els resultats de l'EMEF, amb la finalitat de veure si les dades de telefonia mòbil tenen la consistència i qualitat necessària per a la modelització de la demanda al Projecte VML.

En la segona part dels resultats s'ha realitzat una anàlisi estadística avançada, que és el primer cop que s'aplica en l'àmbit de la demanda de transport. Aquesta anàlisi consisteix en l'Anàlisi de Components Principals (ACP), un exercici de segmentació a partir de l'ACP (utilitzant el k-means) i una Anàlisi de Correspondències Simples (ACS) (Husson, Lê, and Pagès 2011). L'ACP s'aplica a les variables numèriques i serveix per veure les correlacions entre les franges horàries i també quina informació donen les variables categòriques projectades sobre els eixos principals de les variables numèriques. Seguidament s'aplica el clúster amb els individus (són viatges entre parelles OD segons mode/motiu/residència) projectats als eixos principals obtinguts en l'ACP, on agrupa per classes els individus més semblants, però com s'utilitza el k-means el nombre de clústers serà definit prèviament (s'explicarà com s'ha obtingut aquest nombre). I per acabar l'estudi, es realitza l'ACS, per obtenir relacions entre dues zones o agrupacions de zones origen i destinació per valorar diferents els diferents patrons de mobilitat continguts en les dades de manera visual. Totes les anàlisis realitzades es fan amb el software R.

5.1. Anàlisi de Components Principals (ACP)

L'Anàlisi de Components Principals és una eina per visualitzar tota la informació numèrica de les dades. El seu objectiu és representar de forma clara i visual les variacions, relacions, oposicions i altres conceptes de totes les variables numèriques.

La tècnica de l'ACP busca una base de l'hiperespai, anomenada base d'eixos principals, per a projectar-hi tota la informació numèrica. En qualsevol projecció sempre hi ha pèrdua d'informació i per tant, l'objectiu de l'ACP és trobar la base que minimitza aquesta pèrdua d'informació. Finalment, el concepte d'inèrcia és el que quantifica aquesta pèrdua d'informació. La inèrcia de les variables està estretament lligada amb els valors propis d'una certa matriu.

L'Anàlisi de Components Principals (PCA) s'aplica a les taules de dades on les files es consideren individus i les columnes variables quantitatives. Sigui x_{ik} el valor pres per individu i per a la variable k , on varia i de 1 a I i k de 1 a K .

Sigui \bar{x}_k la mitjana de la variable k calculada sobre totes les instàncies individuals de I :

$$\bar{x}_k = \frac{1}{I} \sum_{i=1}^I x_{ik} \quad (5.1)$$

i s_k és la desviació estàndard de la mostra de la variable k (no corregida):

$$s_k = \sqrt{\frac{1}{I} \sum_{i=1}^I (x_{ik} - \bar{x}_k)^2} \quad (5.2)$$

Els passos a seguir són:

1. Estudiar els individus: significa identificar les similituds entre individus des del punt de vista de totes les variables. En altres paraules, proporcionar una tipologia dels individus: quins són els individus més similars (i els més diferents)? Hi ha grups d'individus homogenis en termes de similituds?
2. Estudiar les variables: seguint l'enfocament que es pren per estudiar els individus, també seria possible interpretar les dades de les variables? L'ACP es centra en les relacions lineals entre variables.

Dins de l'estudi dels individus es realitza el següent procediment:

- Núvol dels individus

Un individu és una fila de la taula de dades, és a dir, un conjunt de valors numèrics de K . Els individus evolucionen així dins d'un espai \mathbb{R}^K anomenat "l'espai de l'individu". Si es calcula aquest espai amb la distància euclidiana habitual, la distància entre dos individus i i l s'expressa com:

$$d(i, l) = \sqrt{\sum_{k=1}^K (x_{ik} - x_{lk})^2} \quad (5.3)$$

Aquesta distància serveix per estudiar la forma del núvol de punts o d'individus (denominat N_I).

Si les variables no s'expressen en les mateixes unitats o si no es diuen les unitats de seguretat, és preferible realitzar l'operació d'estandardització, ja que concedeix la mateixa importància a cada variable. Significa que modifica la forma del núvol harmonitzant la seva variabilitat en totes les direccions dels vectors originals (és a dir, les variables K). Llavors x_{ik} passa a ser $(x_{ik} - \bar{x}_k)/s_k$.

- La millor representació al pla N_I

Per obtenir aquesta representació, el núvol N_I es projecta en un pla factorial de \mathbb{R}^K denotat P , escollit de tal manera que es minimitzi la distorsió del núvol de punts. El pla P està seleccionat de manera que les distàncies entre els punts projectades poden ser les més properes possibles a les distàncies entre els punts inicials. Atès que, en la projecció, les distàncies només poden disminuir, s'intenta fer que les distàncies projectades siguin les més altes possibles. En denotar H_i com la projecció de l'individu i en el pla P , el problema consisteix a trobar P , amb:

$$\sum_{i=1}^I OH_i^2 \quad \text{màxim} \quad (5.4)$$

La convenció per a la notació utilitza termes mecànics: O és el centre de gravetat, OH_i és un vector i el criteri és la inèrcia de la projecció de N_I . El criteri que consisteix a augmentar la variància dels punts previstos a un màxim és perfectament apropiat.

Les components de l'ACP s'obtenen a través de la diagonalització de la matriu de correlació que extreu els vectors propis i els valors propis associats. Els vectors propis corresponen als vectors u_s que estan associats amb els valors propis del rang s (denotats λ_s), ja que els valors propis λ_s es classifiquen en ordre descendent. El valor propi λ_s s'interpreta com la inèrcia del núvol N projectada en la component del rang s o, en altres paraules, la variància explicada de la component del rang s .

Per elegir els millors eixos principals es pot fer per dos criteris: la suma de la inèrcia total sigui superior al 80% de la informació total; o utilitzant el criteri de Kaiser, que consisteix a eliminar totes les components amb valors propis més petits que 1, sent aquest valor propi igual a la informació comptabilitzada per un sol element.

L'estudi de les variables segueix el següent procediment:

- Núvol de les variables

Ara la taula de dades s'observa com un conjunt de columnes. Una variable és una de les columnes de la taula, és a dir, un conjunt de valors numèrics que es representa per un punt de l'espai vectorial amb dimensions I , denotat \mathbb{R}^I (i conegut com l'espai de les variables). El vector connecta l'origen de \mathbb{R}^I al punt. Tots aquests vectors constitueixen el núvol de variables i aquest conjunt es denota N_K .

El producte escalar entre dues variables k i l s'expressa de la manera següent:

$$\sum_{i=1}^I x_{ik} \times x_{il} = \|k\| \times \|l\| \times \cos(\theta_{kl}) \quad (5.5)$$

Les variables utilitzades normalment estan centrades, llavors, la norma per a una variable és igual a la seva desviació estàndard multiplicada per l'arrel quadrada de I , i el producte escalar s'expressa de la següent manera:

$$\sum_{i=1}^I (x_{ik} - \bar{x}_k) \times (x_{il} - \bar{x}_l) = I \times s_k \times s_l \times \cos(\theta_{kl}) \quad (5.6)$$

a la part dreta de l'equació, es pot identificar la covariància entre les variables k i l .

De la mateixa manera, dividint cada terme en l'equació per les desviacions estàndard s_k i s_l de les variables k i l , s'obté la següent relació:

$$r(k, l) = \cos(\theta_{kl}) \quad (5.7)$$

Per a la construcció del núvol de les variables, el criteri utilitzat és que les variables v_s maximitzin el criteri $\sum_{k=1}^K (OH_k^s)^2$. Atès que la projecció d'una variable k és igual al cosinus d'angle θ_k^s , el criteri és maximitzar:

$$\sum_{k=1}^K \cos^2 \theta_k^s = \sum_{k=1}^K r^2(k, v_s) \quad (5.8)$$

L'expressió anterior mostra que v_s és la nova variable, que més fortament està correlacionada amb totes les variables inicials K (amb la restricció d'ortogonalitat de v_t ja trobada). Com a resultat, es pot dir que v_s es tracta d'una variable sintètica.

5.2. Clustering

El terme clustering fa referència a un ampli ventall de tècniques no supervisades que té com a finalitat trobar patrons o grups (clústers) dins d'un conjunt d'observacions. Les particions s'estableixen de manera que, les observacions que estan dins d'un mateix grup, són similars entre elles i diferents de les observacions d'altres grups. Es tracta d'un mètode no supervisat, ja que el procés ignora la variable resposta que indica a quin grup pertany realment cada observació (si és que existeix tal variable). Aquesta característica diferencia al clustering de les tècniques estadístiques conegudes com a anàlisi discriminant, que entren un set d'entrenament en el qual es coneix la veritable classificació.

Donada la popularitat del clustering en disciplines molt diferents, s'ha desenvolupat multitud de variants i adaptacions dels seus mètodes i algoritmes. Es poden diferenciar dos grans grups:

- *Partitioning Clustering*: Aquest tipus d'algorismes requereixen que l'usuari especifiqui per endavant el nombre de clústers que es crearan (*K-means* i *K-medoids*).
- *Hierarchical Clustering*: Aquest tipus d'algorismes no requereixen que l'usuari especifiqui per endavant el nombre de clústers. (*Agglomerative clustering*, *divisive clustering*).

En aquest treball s'utilitzarà el partitioning clustering, concretament K-means. Ja que es té una base de dades molt gran i és l'únic mètode que es pot utilitzar amb totes les observacions. En canvi, els altres és necessari agafa una mostra aleatòria poc representativa per utilitzar-los.

5.2.1. K-means clustering

El mètode K-means clustering (MacQueen, 1967) agrupa les observacions en K clústers diferents, on el nombre K ho determina l'analista. K-means troba els K millors clústers, entenent com a millor clúster aquell on la variància interna (variància intra-clúster) sigui la més petita possible. Es tracta per tant d'un problema d'optimització, en el qual es reparteixen les observacions en k clústers de manera que la suma de les variàncies internes de tots ells sigui la menor possible. Per poder solucionar aquest problema cal definir una manera de quantificar la variància interna.

Es considera C_1, \dots, C_K com els sets formats pels índexs de les observacions de cada un dels clústers. La nomenclatura emprada per indicar que l'observació i pertany al clúster k és: $i \in C_k$. Tots els sets satisfan dues propietats:

1. $C_1 \cup C_2 \cup \dots \cup C_K = \{1, \dots, n\}$. Significa que tota observació pertany almenys a un dels k clústers.
2. $C_k \cap C_{k'} = \emptyset, \forall k \neq k'$ implica que els clústers no es superposen, cap observació pertany a més d'un clúster alhora.

Dues de les mesures més comunament emprades defineixen la variància interna d'un clúster ($W(C_k)$) com:

- La suma de les distàncies euclidianes al quadrat entre cada observació (x_i) i el centroid (μ) del seu clúster. Això equival a la suma de quadrats interns del clúster:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

- La suma de les distàncies euclidianes al quadrat entre tots els parells d'observacions que formen el clúster, dividida entre el nombre d'observacions del clúster:

$$W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2$$

Minimitzar la suma total de la variància interna $\sum_{k=1}^K W(C_k)$ de forma exacta és un procés molt complex a causa de la immensa quantitat de formes en què n observacions es poden dividir en k grups.

Una forma d'implementar l'algoritme de K-means clustering és:

1. Especificar el nombre K de clústers que es volen crear.
2. Seleccionar de forma aleatòria k observacions del set de dades com centroides inicials.
3. Assignar cadascuna de les observacions al centroide més proper.
4. Per a cada un dels K clústers re-calcular el centroide.
5. Repetir els passos 3 i 4 fins que les assignacions no canviïn o s'arribi al nombre màxim d'iteracions establert.

A causa que l'algoritme del K-means no avalua totes les possibles distribucions de les observacions sinó només part d'elles, els resultats obtinguts depenen de l'assignació aleatòria inicial (pas 1). Per aquesta raó és important executar l'algoritme diverses vegades (20-50), cadascuna amb una assignació aleatòria inicial diferent, i seleccionar aquella que hagi aconseguit un menor valor de variància total.

5.3. Anàlisi de Correspondències Simples (ACS)

L'Anàlisi de Correspondències Simples (ACS) és una tècnica exploratòria amb l'objectiu de contrastar visualment dues variables categòriques amb la menor pèrdua d'informació possible. Amb aquesta tècnica es pretén trobar relacions i oposicions entre dues variables categòriques de la base de dades. També resulta interessant trobar factors latents associats a les variables.

Primerament, es construeix una taula de contingència (Figura 5.1) de la següent manera: per a n individus, els valors estan disponibles per a dues variables categòriques V_1 (amb I categories o nivells) i V_2 (amb J categories). La taula de contingència porta el terme general x_{ij} com el nombre d'individus que figuren en la categoria i de V_1 i j de V_2 .

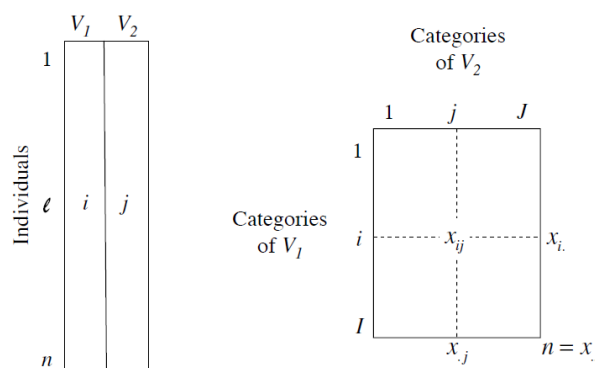


Figura 5.1: Construcció d'una taula de contingència. Font: (Husson, Lê, and Pagès 2011)

La Figura 5.1 descriu les indicacions generals per a una taula de contingència amb les dues variables categòriques V_1 i V_2 definides per a n individus; l'individu porta les categories i (per V_1) i j (per a V_2): aquestes es comptabilitzen en x_{ij} .

Les sumes de la taula es denoten substituint l'índex de x_{ij} en què es fa el càlcul per un punt. Així:

$$x_{i\bullet} = \sum_{j=1}^J x_{ij} \quad x_{\bullet j} = \sum_{i=1}^I x_{ij} \quad n = x_{\bullet\bullet} = \sum_{i,j} x_{ij} \quad (5.9)$$

Finalment, en l'anàlisi de correspondència, es consideren les taules de probabilitat associades a taules de contingència com el terme general $f_{ij} = x_{ij}/n$, la probabilitat de portar ambdues categories i (de V_1) i les de j (V_2). Les sumes d'aquesta taula, que també es coneixen com a probabilitats marginals, són desglossades per:

$$f_{i\bullet} = \sum_{j=1}^J f_{ij} \quad f_{\bullet j} = \sum_{i=1}^I f_{ij} \quad f_{\bullet\bullet} = \sum_{i,j} f_{ij} = 1 \quad (5.10)$$

El principi de la construcció d'una taula de contingència implica que l'objectiu d'estudiar una taula com aquesta és examinar les relacions entre les respostes a dues preguntes. Cal assenyalar aquí, que com a la majoria de les taules utilitzades a l'ACS, s'està segur que existeix una relació. A la llum del significat d'aquestes preguntes, es realitza una prova i si no existeix cap relació entre les preguntes, seria una mica sorprenent, i es qüestionaria la qualitat de les dades.

La prova s'utilitza per comprovar la importància de la desviació global de la mostra real del model d'independència. S'expressa de la manera següent:

$$\chi^2 = \sum_{i,j} \frac{(\text{Tamany mostra real} - \text{Tamany mostra teòrica})^2}{\text{Tamany mostra teòrica}} \quad (5.11)$$

$$\chi^2 = \sum_{i,j} \frac{(nf_{ij} - nf_{i\bullet}f_{\bullet j})^2}{nf_{i\bullet}f_{\bullet j}} = n \sum_{i,j} \frac{(f_{ij} - f_{i\bullet}f_{\bullet j})^2}{f_{i\bullet}f_{\bullet j}} = n\Phi^2 \quad (5.12)$$

Llavors el contrast d'hipòtesi per realitzar aquest test és:

- $H_0: V_1$ i V_2 són independents
- $H_1: V_1$ i V_2 no són independents

Per tant si $\chi^2 < \chi^2_{\alpha,n}$ no es rebutja la hipòtesi nul·la (H_0), i es pot dir que les dues variables estudiades són independents.

- Núvol de perfils fila

A partir de la taula de perfils de files, es construeix un nucli de punts denotats N_I , dins de l'espai \mathbb{R}^J en el qual cada dimensió correspon a una categoria de la variable V_2 . Aquesta construcció és similar a la dels núvols d'individus en l'ACP. Cada fila i , té un punt corresponent a les coordenades de la dimensió j que és $f_{ij} = f_{i\bullet}$. Aquest núvol es complementa amb el punt mitjà G_I , amb un valor de coordenada j^{th} de f_j (Figura 5.2).

A més de les transformacions en perfils fila, en comparació amb el núvol d'individus en l'ACP, el núvol de les files a l'ACS té les dues característiques següents:

1. Cada punt i s'atribueix a un pes $f_{i\bullet}$; aquest pes s'imposa i representa una part integral de l'ACS. Per a un determinat perfil, s'atribueix una influència a cada categoria que augmenta amb la seva freqüència. Amb aquest pes, la mitjana del perfil G_I és el centre de gravetat de N_I . Aquest punt G_I es pren com a origen dels eixos (pel que fa als individus en l'ACP).
2. La distància atribuïda a l'espai \mathbb{R}^J consisteix a atribuir el pes $1/f_{\bullet j}$ a la dimensió j . El quadrat de la distància (anomenada distància χ^2) entre els punts i i l s'expressa com:

$$d_{\chi^2}^2(i, l) = \sum_{j=1}^J \frac{1}{f_{\bullet j}} \left(\frac{f_{ij}}{f_{i\bullet}} - \frac{f_{lj}}{f_{l\bullet}} \right)^2 \quad (5.13)$$

La principal justificació per aquesta distància és indirecta i es troba en la següent propietat. Al considerar el pes $f_{i\bullet}$, la inèrcia del punt i respecte a la G_I s'expressa com:

$$Inertia\left(\frac{i}{G_I}\right) = f_{i\bullet} \cdot d_{\chi^2}^2(i, G_I) = f_{i\bullet} \cdot \sum_{j=1}^J \frac{1}{f_{\bullet j}} \left(\frac{f_{ij}}{f_{i\bullet}} - f_{\bullet j} \right)^2 \quad (5.14)$$

$$Inertia\left(\frac{i}{G_I}\right) = \sum_{j=1}^J \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})^2}{f_{i\bullet} f_{\bullet j}} \quad (5.15)$$

Fins al factor multiplicatiu n , es reconeix l'aportació de la fila i de la prova χ^2 . El resultat és que la inèrcia total del núvol de punts N_I respecte a G_I és igual (fins al factor multiplicatiu n) al criteri χ^2 . També es pot dir que aquesta inèrcia és igual a Φ^2 . Examinar la distribució de N_I en termes de G_I significa examinar la diferència entre les dades i el model independent. Això és el que fa l'ACS en ressaltar les indicacions de major inèrcia per a N_I .

Aquesta inèrcia, igual a Φ^2 , ofereix informació vital, ja que mesura la intensitat de la relació entre les dues variables de la taula de contingència. Això prové de l'ACP estandarditzada en la qual la inèrcia total és igual a la quantitat de variables i no depèn de les mateixes dades.

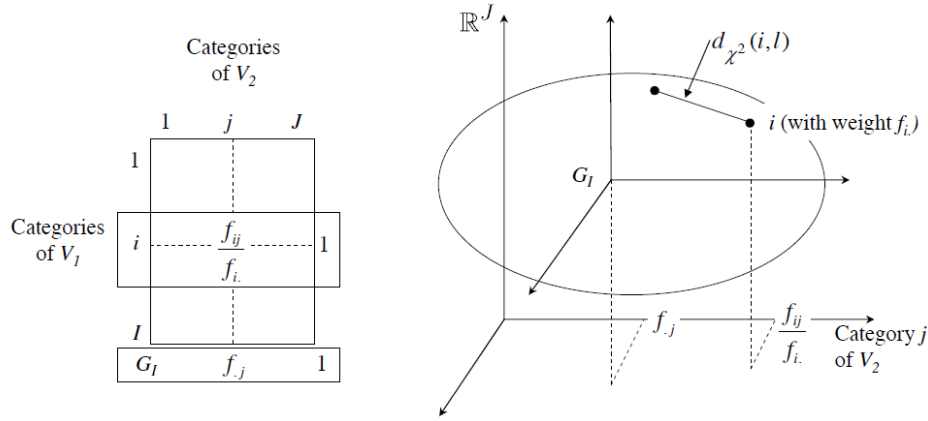


Figura 5.2: Núvol de perfils fila a l'ACS. Font: (Husson, Lê, and Pagès 2011)

- Núvol de perfils columna

En les taules de contingència, les files i les columnes tenen funcions simètriques: es poden estudiar $V_1 \times V_2$ o $V_2 \times V_1$. Aquesta és una de les principals diferències entre l'ACS i l'ACP en què les files (individus) i les columnes (variables) no s'analitzen de la mateixa manera: es consideren distàncies entre individus i correlacions entre variables. En conseqüència, l'ACS construeix el núvol de perfils columna de la mateixa forma que el núvol de perfils fila (Figura 5.3):

1. Es consideren perfils columna depenent de si està interessat en les files o les columnes, no s'analitzarà la mateixa taula; $f_{ij}/f_{i.}$ en un cas, $f_{ij}/f_{.j}$ en l'altre. Aquesta és una de les principals deficiències entre ACS i ACP, en què la mateixa transformació de les dades s'utilitza per estudiar individus i variables.
2. Cada columna té un punt corresponent a \mathbb{R}^I , de la qual la coordenada en la dimensió i és $f_{ij}/f_{.j}$. Aquests punts conformen el núvol N_J .
3. Cada punt j s'atribueix a un pes $f_{.j}$. Amb aquest pes, el centre de gravetat del núvol, denominat G_J , és igual a la mitjana del perfil. L'origen dels eixos es troba a G_J .

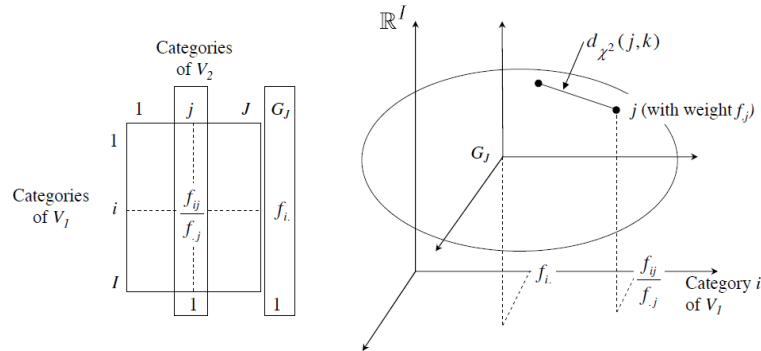


Figura 5.3: Núvol de perfils columna a l'ACS. Font: (Husson, Lê, and Pagès 2011)

A \mathbb{R}^I , la distància atribueix el pes $1/f_{i.}$ a la dimensió i . La distància (quadrada) entre dues columnes j i k s'expressa com:

$$d_{\chi^2}^2(j, k) = \sum_{i=1}^I \frac{1}{f_{i\cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - \frac{f_{ik}}{f_{\cdot k}} \right)^2 \quad (5.16)$$

La inèrcia a la columna j respecte al punt G_j s'expressa com:

$$\begin{aligned} Inertia\left(\frac{i}{G_j}\right) &= f_{\cdot j} d_{\chi^2}^2(j, G_j) \\ &= f_{\cdot j} \sum_{i=1}^I \frac{1}{f_{i\cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - f_{i\cdot} \right)^2 = \sum_{j=1}^J \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}} \end{aligned} \quad (5.17)$$

Es reconeix la contribució (fins a un factor multiplicatiu n) de la columna j a χ^2 . La inèrcia total de N_j és, per tant, la mateixa que la de N_I ($= \frac{1}{n} \chi^2$): estudiar la distribució de N_j al voltant de G_j significa examinar la relació entre ambdues variables V_1 i V_2 .

- Núvols apropiats per a N_I i N_j

Es treballa i es segueix el mateix procediment que en el núvol d'individus en l'ACP.

6. Descripció del tractament de les dades

Les dades originals subministrades per Kineo s'han descrit en 3.2.1, a partir d'aquestes matrius s'ha creat una matriu general, una per cadascun dels motius de viatge i una per a cadascun dels llocs de residència. S'ha de mencionar, que els viatges calculats segons el mode del transport s'han obtingut després d'un procés de fusió amb dades prèvies descrites a (Montero et al. 2018) i no forma part del present treball. A continuació, s'explica el procés que s'ha dut a terme des de les matrius proporcionades per Kineo fins a obtenir les matrius en què es realitza l'anàlisi.

Primerament, cadascuna de les matrius s'agrupa per zona d'origen, zona de destinació i per franja horària, i es combinen per crear una matriu amb tots els registres. A part, es crea l'estructura de com es vol la matriu per fer l'anàlisi posterior, o sigui, el nom i l'ordre de cadascuna de les variables. Un cop es tenen les dues matrius es filtren per obtenir una matriu general, una agrupada segons el motiu del viatge i una altra segons el lloc de residència.

Cadascuna de les matrius contenen 48 variables i 388129 observacions. Hi ha 36 variables numèriques i 12 de tipus categòric.

6.1. Variables de tipus numèric

Tot seguit es descriuen les variables de tipus numèric:

- id: identificador
- ori: origen
- des: destinació
- od: parell origen-destinació (T_{ij})
- tot_emit: total viatges segons l'enquesta emit de la parella OD
- tot_kineo: total viatges segons kineo de la parella OD
- h_t _kineo: viatges realitzats entre t i $t+1$, $t \in \{0, \dots, 23\}$. Per exemple: h1_kineo són viatges realitzats entre la 1 i les 2 (un total de 24 variables)
- pobori: població de la zona de transport origen
- pobdes: població de la zona de transport destinació

- dm_kineo_car: total de viatges mig amb cotxe segons Kineo
- dm_kineo_bus: total de viatges mig amb bus segons Kineo
- dm_kineo_met: total de viatges mig amb metro segons Kineo
- dm_kineo_other: total de viatges mig amb altres transports segons Kineo

6.2. Variables categòriques

A continuació, es descriuen les variables de tipus numèric:

- munori: municipi d'origen
 - Badalona
 - Barcelona
 - Castelldefels
 - Cornellà de Llobregat
 - Esplugues de Llobregat
 - Gavà
 - L'HospitaletLl (l'Hospitalet de Llobregat)
 - Montcada i Reixac
 - Montgat
 - Prat de Llobregat, el (el Prat del Llobregat)
 - Sant Adrià de Besós
 - Sant Boi de Llobregat
 - Sant Feliu de Llobregat
 - Sant Joan Despí (Sant Joan Despí)
 - Sant Just Desvern
 - Santa Coloma de Gramenet
 - Tiana
 - Viladecans
 - Gate: municipis exteriors
- mundes: municipi de destinació; són els mateixos municipis que la variable munori.
- mundisori: codi del districte on pertany el municipi d'origen, on les primeres 6 xifres fan referència al codi del municipi i les altres dos al districte pertanyen al municipi. Els codis són:
 - 08015501: Badalona districte 1
 - 08015502: Badalona districte 2
 - 08015503: Badalona districte 3
 - 08015504: Badalona districte 4
 - 08015505: Badalona districte 5
 - 08015506: Badalona districte 6
 - 08015507: Badalona districte 7
 - 08015508: Badalona districte 8

- 08015509: Badalona districte 9
- 08019301: Barcelona districte 1
- 08019302: Barcelona districte 2
- 08019303: Barcelona districte 3
- 08019304: Barcelona districte 4
- 08019305: Barcelona districte 5
- 08019306: Barcelona districte 6
- 08019307: Barcelona districte 7
- 08019308: Barcelona districte 8
- 08019309: Barcelona districte 9
- 08019310: Barcelona districte 10
- 08019311: Barcelona districte 11
- 08056901: Castelldefels districte 1
- 08056902: Castelldefels districte 2
- 08056903: Castelldefels districte 3
- 08056904: Castelldefels districte 4
- 08056905: Castelldefels districte 5
- 08073401: Cornellà de Llobregat districte 1
- 08073402: Cornellà de Llobregat districte 2
- 08073403: Cornellà de Llobregat districte 3
- 08073404: Cornellà de Llobregat districte 4
- 08073405: Cornellà de Llobregat districte 5
- 08073406: Cornellà de Llobregat districte 6
- 08077101: Esplugues de Llobregat districte 1
- 08077102: Esplugues de Llobregat districte 2
- 08077103: Esplugues de Llobregat districte 3
- 08077108: Esplugues de Llobregat districte 8
- 08077109: Esplugues de Llobregat districte 9
- 08077110: Esplugues de Llobregat districte 10
- 08077111: Esplugues de Llobregat districte 11
- 08089801: Gavà districte 1
- 08089802: Gavà districte 2
- 08089803: Gavà districte 3
- 08089804: Gavà districte 4
- 08101701: L'Hospitalet de Llobregat districte 1
- 08101702: L'Hospitalet de Llobregat districte 2
- 08101703: L'Hospitalet de Llobregat districte 3
- 08101704: L'Hospitalet de Llobregat districte 4
- 08101705: L'Hospitalet de Llobregat districte 5
- 08101706: L'Hospitalet de Llobregat districte 6
- 08125201: Montcada i Reixac districte 1
- 08125202: Montcada i Reixac districte 2
- 08125203: Montcada i Reixac districte 3
- 08125204: Montcada i Reixac districte 4

- 08125205: Montcada i Reixac districte 5
- 08125206: Montcada i Reixac districte 6
- 08126501: Montgat districte 1
- 08169101: el Prat de Llobregat districte 1
- 08169102: el Prat de Llobregat districte 2
- 08169103: el Prat de Llobregat districte 3
- 08169104: el Prat de Llobregat districte 4
- 08169105: el Prat de Llobregat districte 5
- 08194401: Sant Adrià de Besòs districte 1
- 08194402: Sant Adrià de Besòs districte 1
- 08194403: Sant Adrià de Besòs districte 1
- 08194404: Sant Adrià de Besòs districte 1
- 08194405: Sant Adrià de Besòs districte 1
- 08194406: Sant Adrià de Besòs districte 1
- 08200901: Sant Boi de Llobregat districte 1
- 08200902: Sant Boi de Llobregat districte 2
- 08200903: Sant Boi de Llobregat districte 3
- 08200904: Sant Boi de Llobregat districte 4
- 08200905: Sant Boi de Llobregat districte 5
- 08200906: Sant Boi de Llobregat districte 6
- 08211401: Sant Feliu de Llobregat districte 1
- 08211402: Sant Feliu de Llobregat districte 2
- 08211403: Sant Feliu de Llobregat districte 3
- 08211404: Sant Feliu de Llobregat districte 4
- 08211405: Sant Feliu de Llobregat districte 5
- 08211406: Sant Feliu de Llobregat districte 6
- 08211407: Sant Feliu de Llobregat districte 7
- 08217201: Sant Joan Despí districte 1
- 08217202: Sant Joan Despí districte 2
- 08217203: Sant Joan Despí districte 3
- 08217204: Sant Joan Despí districte 4
- 08217205: Sant Joan Despí districte 5
- 08221201: Sant Just Desvern districte 1
- 08245701: Sant Just Desvern districte 2
- 08245702: Sant Just Desvern districte 3
- 08245703: Sant Just Desvern districte 4
- 08245704: Sant Just Desvern districte 5
- 08245705: Sant Just Desvern districte 6
- 08245706: Sant Just Desvern districte 7
- 08282401: Tiana districte 1
- 08301501: Viladecans districte 1
- 08301502: Viladecans districte 2
- gate11: districtes exteriors

- mundisdes: és la mateixa codificació que la variable mundisori, però en aquest cas és el districte del municipi de destinació.
- nombarori: nom del barri d'origen
 - Baró de Viver
 - Can Baró
 - Can Peguera
 - Canyelles
 - Ciutat Meridiana
 - Diagonal Mar i el Front Marítim del Poblenou
 - el Baix Guinardó
 - el Barri Gòtic
 - el Besòs i el Maresme
 - el Bon Pastor
 - el Camp d'en Grassot i Gràcia Nova
 - el Camp de l'Arpa del Clot
 - el Carmel
 - el Clot
 - el Coll
 - el Congrés i els Indians
 - el Fort Pienc
 - el Guinardó
 - el Parc i la Llacuna del Poblenou
 - el Poble Sec
 - el Poblenou
 - el Putxet i el Farró
 - el Raval
 - el Turó de la Peira
 - Horta
 - Hostafrancs
 - l'Antiga Esquerra de l'Eixample
 - la Barceloneta
 - la Bordeta
 - la Clota
 - la Dreta de l'Eixample
 - la Font d'en Fargues
 - la Font de la Guatlla
 - la Guineueta
 - la Marina de Port
 - la Marina del Prat Vermell
 - la Maternitat i Sant Ramon
 - la Nova Esquerra de l'Eixample
 - la Prosperitat
 - la Sagrada Família
 - la Sagrera

- la Salut
 - la Teixonera
 - la Trinitat Nova
 - la Trinitat Vella
 - la Vall d'Hebron
 - la Verneda i la Pau
 - la Vila de Gràcia
 - la Vila Olímpica del Poblenou
 - les Corts
 - les Roquetes
 - les Tres Torres
 - Montbau
 - Navas
 - Pedralbes
 - Porta
 - Provençals del Poblenou
 - Sant Andreu
 - Sant Antoni
 - Sant Gens dels Agudells
 - Sant Gervasi - Galvany
 - Sant Gervasi - la Bonanova
 - Sant Mart de Provençals
 - Sant Pere, Santa Caterina i la Ribera
 - Sants
 - Sants - Badal
 - Sarrià
 - Torre Baró
 - Vallcarca i els Penitents
 - Vallvidrera, el Tibidabo i les Planes
 - Verdun
 - Vilapicina i la Torre Llobeta
- nombardes: nom del barri de destinació; són els mateixos barris que té la variable nombarori.
 - comori: comarca de la zona d'origen
 - Baix Llobregat
 - Barcelonès
 - Maresme
 - Vallès Occidental
 - comdes: comarca de la zona de destinació; les mateixes comarques que comori.
 - munos: codi del municipi d'origen, és el mateix ordre de municipis que en la variable munori
 - O_Bad

- O_BCN
 - O_Cas
 - O_Cor
 - O_Esp
 - O_Gava
 - O_LHo
 - O_Mnc
 - O_Mng
 - O_Prats
 - O_SAB
 - O_SBL
 - O_SCo
 - O_SFL
 - O_SJoD
 - O_SJuD
 - O_Tia
 - O_Vil
 - O_Gate
- munds: codi del municipi de destinació; és el mateix codi que la variable munori, però canvia el codi O_ on en aquesta variable és D_.
 - purpose: motiu del viatge:
 - Feina: el motiu del viatge realitzat és anar a la feina.
 - Casa: la utilització del transport és per tornar a casa.
 - Freqüent: en aquest cas no es pot posar una etiqueta específica, però va amb molta freqüència (escola, hospital, gimnàs, ...).
 - PocFreqüent: aquesta opció és cap de les anteriors, o sigui, és un lloc on va amb molt poca freqüència.
 - resident: d'on és resident la persona que realitza el viatge:
 - PC: Primera Corona de Barcelona, no inclou Barcelona.
 - BCN: Barcelona.
 - RMB: Regió Metropolitana de Barcelona, no inclou cap de les anteriors.

6.2.1. Selecció de les variables

Hi ha 57 variables i es seleccionen 46, perquè són les variables adequades per els indicadors globals i per mode, motiu i lloc de residència que es busca valorar i comparar amb EMEF en l'exercici de validació. Les variables seleccionades són: id, ori, des, od, tot_emit, tot_kineo, franges horàries (forma h_t -kineo), munori, munds, mundisori, mundisdes, nombarori, nombardes, pobori, pobdes, comori, comdes, munos, munds, dm_kineo_car, dm_kineo_bus, dm_kineo_met, dm_kineo_other, purpose i resident.

7. Resultats

7.1. Preprocessament de les dades

El preprocessament de les dades es realitza amb l'objectiu de cercar valors anòmals, incoherències a les dades i missings.

No s'han observat errors a la base de dades. L'única cosa que s'ha modificat és: les variables que presenten NA's (nombarori, nombardes, comoroi, comdes, munos i munds) se'ls ha de modificar aquesta etiqueta, perquè aquestes observacions són viatges que provenen de l'exterior (per exemple, nombarori, és un viatge que té origen a un lloc que no pertany a cap barri de Barcelona), i la nova etiqueta és per a cadascuna de les parelles de variables: Barri_Gate (nombarori i nombardes), Com_Gate (comori i comdes), O_Gate (munos) i D_Gate (munds). On la terminologia "gate" té un significat de porta exterior.

7.2. Dades de mobilitat a grans àmbits a la Regió Metropolitana de Barcelona (RMB)

Per començar l'estudi de la mobilitat a la RMB es realitza una anàlisi territorial per veure els tipus de fluxos que existeixen. Cal destacar que les dades que no tenen a veure amb la mobilitat a la PC (Primera Corona) no estan ben quantificades, ja que no és l'objectiu principal d'estudi del Projecte VML. Per això, els resultats que es refereixen a la RMB que no inclou la PC no són massa fiables alhora de la seva interpretació. També no es tenen en compte els residents de fora de la RMB, per exemple un turista resident a una ciutat europea que ve a visitar Barcelona, aquest tipus de persona no es té en compte en les dades de Kineo.

A la RMB es realitzen 13,5 milions de viatges al dia de mitjana en dia feiner (Taula 7.1) considerant tots els motius i tots els modes de transport per persones residents a tota la RMB majors de 16 anys. Aquesta Regió s'ha dividit en 3 subàrees per veure millor els tipus de fluxos, aquestes són: viatges dins de la Primera Corona (IntraPC), viatges dins RMB (RMBnoPC) exclouent la PC i els viatges de connexió d'aquestes dues zones (*commuters*). Com s'observa a la Figura 7.1 la quantitat més gran de viatges realitzats es

fan a la Primera Corona amb un 58% dels viatges totals i els viatges de connexió només un 10%.

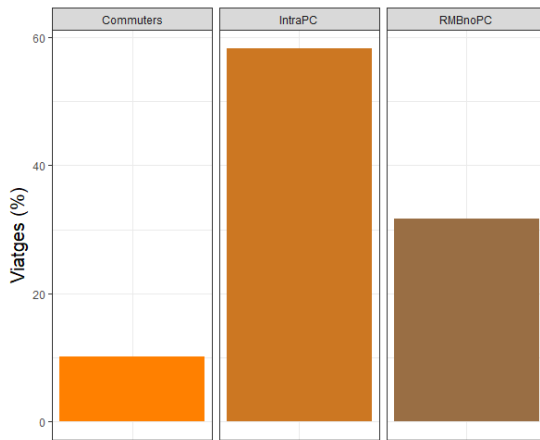


Figura 7.1: Àrea de mobilitat RMB

Taula 7.1: Àrea de mobilitat RMB

Zona	Viatges	Percentatge
Commuters	1.370.789	10,11%
RMBnoPC	4.293.043	31,65%
IntraPC	7.898.975	58,24%
Total	13.562.808	100%

Observant els resultats de l'EMEF, on a la PC es realitzen 9,5 milions de desplaçaments, i amb les dades que s'està treballant hi ha 7,9 milions, aquesta diferència es deguda a factors com els que s'han comentat que hi ha viatges que no es tenen en compte, i que l'EMEF utilitza un factor de correcció que no s'explica per traslladar les dades de població major de 15 anys al total de la població. Passa el mateix amb els viatges totals, segons l'EMEF són 18,8 milions de desplaçaments a la RMB i amb les dades de Kineo són 13,5 milions.

Per veure millor les característiques dels fluxos, les dades es desagreguen segons el motiu de viatge i segons el lloc de residència.

Els viatges segons el motiu a la RMB es distribueixen de manera similar per cadascun dels fluxos considerats, malgrat que dintre de la Primera Corona és on hi ha una quantitat de viatges més elevada per a cada motiu (Taula 7.2). Tant la taula (Taula 7.3) com el gràfic (Figura 7.2), el tant per cent dels viatges és per motiu.

Motiu	RMBnoPC	IntraPC	Commuters	Total
Freqüent	2.139.477	4.285.059	716.742	7.141.278
Casa	1.515.527	2.460.051	436.802	4.412.380
Feina	599.899	1.048.950	196.035	1.844.884
Poc Freqüent	38.141	104.915	21.211	164.267
Total	4.293.043	7.898.975	1.370.789	13.562.808

Taula 7.2: Àrea de mobilitat segons el motiu del viatge RMB

Motiu	RMBnoPC	IntraPC	Commuters	Total
Freqüent	29,96%	60,00%	10,04%	100%
Casa	34,35%	55,75%	9,90%	100%
Feina	32,52%	56,86%	10,63%	100%
Poc Freqüent	23,22%	63,87%	12,91%	100%

Taula 7.3: Àrea de mobilitat segons el motiu del viatge RMB (%)

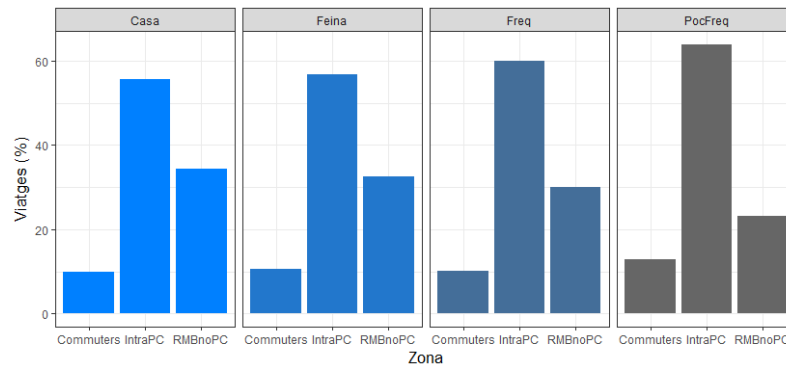


Figura 7.2: Àrea de mobilitat segons el motiu del viatge RMB

En aquest cas és més difícil de comparar amb l'EMEF, donat que la definició dels motius no és la mateixa per KINEO, que per l'estudi EMEF. En les dades de Kineo el motiu anar a la feina, només fa referència a qui va a treballar, en canvi, a l'EMEF (la seva terminologia és mobilitat ocupacional) hi pertanyen els qui van a treballar i a estudiar. Els qui van a estudiar en aquest treball estan inclosos al motiu desplaçament freqüent.

En els fluxos de viatges considerats, si es desagrega pel lloc de residència del viatger, s'observa que els residents a RMB (no inclou PC) són els que realitzen més viatges (Taula 7.4), però només amb els residents a Barcelona es té un 35% (aproximadament) dels viatges totals. Seguidament, es comenten les característiques més rellevants (Taula 7.5 i Figura 7.3):

- Els residents a la ciutat de Barcelona realitzen el 91% dels seus viatges dintre de la Primera Corona.
- Els residents a la Primera Corona (no inclou Barcelona) fan un 88% dels viatges a dintre de la Primera Corona.
- Els residents a la Regió Metropolitana de Barcelona (no inclou Primera Corona) realitzen un 77% dels viatges dintre a aquesta regió i un 15% dels viatges són de connexió entre aquesta zona i la Primera Corona.

Residència	RMBnoPC	IntraPC	Commuters	Total
BCN	135.650	4.380.139	274.378	4.790.167
PC	109.970	3.083.896	313.200	3.507.066
RMB	4.047.423	434.940	783.211	5.265.574
Total	4.293.043	7.898.975	1.370.789	13.562.807

Taula 7.4: Àrea de mobilitat segons el lloc de residència RMB

Residència	RMBnoPC	IntraPC	Commuters	Total
BCN	2,83%	91,44%	5,73%	100%
PC	3,14%	87,93%	8,93%	100%
RMB	76,87%	8,26%	14,87%	100%

Taula 7.5: Àrea de mobilitat segons el lloc de residència RMB (%)

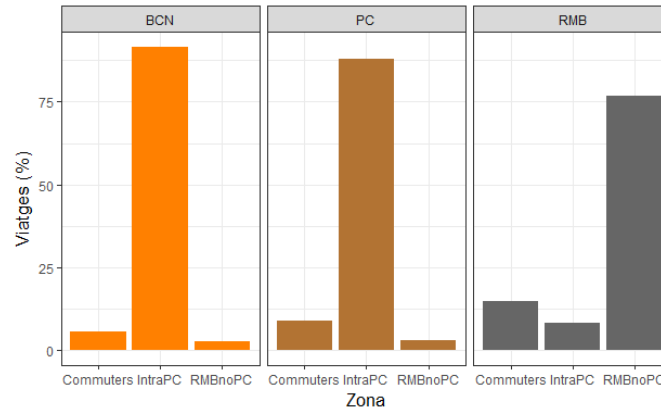


Figura 7.3: Àrea de mobilitat segons el lloc de residència RMB

A continuació s'analitza en quin mode de transport es realitzen aquests viatges, per tant les dades anteriors de viatges es segmenten segons 4 grans modes: cotxe (fa referència a tot el transport privat motoritzat), autobús (i mitjans de transport relacionats), metro (també inclou metro, tramvia, ferrocarrils, trens...) i altres (transport no motoritzat com ara la bicicleta, caminar...). En general, el repartiment modal a la RMB té les següents característiques (Figura 7.4 i Taula 7.6):

- El mode de transport més utilitzat és la mobilitat motoritzada, amb un 56% dels viatges diaris totals.
- Pel que fa als viatges en transport públic, predominen els modes ferris (9%), seguidament de l'ús de l'autobús (7%).
- El cotxe, amb un 24% dels desplaçaments, és el segon mode de transport més utilitzat.

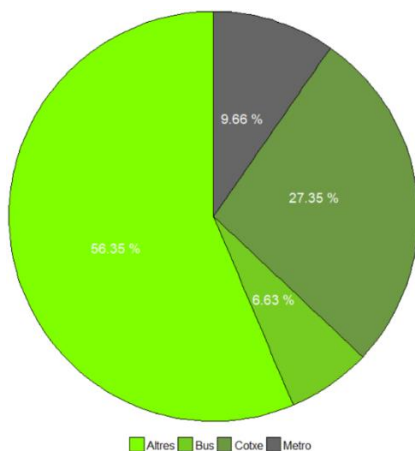


Figura 7.4: Mode de transport RMB

Mode de transport	Totals	Viatges	Percentatge
Cotxe	Transport privat	3.709.974	27,36%
Autobús	Transport públic	899.307	6,63%
Metro		1.310.407	9,66%
Altres	Transport no motoritzat	7.643.117	56,35%
	Total	13.562.805	100%

Taula 7.6: Mode de transport RMB

En aquest cas, els resultats són semblants als de l'EMEF, sobretot els que fan referència al transport públic: 16,9% (EMEF) i 16,3% (Kineo). En els altres modes hi ha diferències petites: en el transport privat són un 35% (EMEF) i un 27% (Kineo) del total de desplaçaments i en el transport no motoritzat un 48% (EMEF) i un 56% (Kineo) del total de viatges.

Igual que s'ha fet abans, les dades de viatges segons el mode de transport es desagreguen segons el motiu de viatge i segons el lloc de residència.

L'ús dels modes de transport segons el motiu del viatge indica:

- En la mobilitat per motius freqüent o poc freqüent, les tornades a casa i la mobilitat ocupacional, predomina l'ús dels models no motoritzats (altres) (Taula 7.8). Aquesta característica és deguda al procés d'elaboració de la mobilitat modal, per fusió amb les dades de l'EMIT i revela algunes discrepàncies amb els resultats de l'enquesta EMEF.
- Per a cadascun dels motius, els viatges en cotxe són aproximadament quatre vegades més que els de l'autobús (Taula 7.7).
- Els desplaçaments amb motiu poc freqüent són modalment els més diferents dels altres, ja que el total de viatges realitzats segons aquest motiu utilitza menys el transport no motoritzat i més el motoritzat (Figura 7.5).

Motiu	Cotxe	Autobús	Metro	Altres	Total
Freqüent	1.967.584	496.341	713.634	3.963.716	7.141.275
Casa	1.182.005	271.979	398.582	2.559.814	4.412.380
Feina	508.900	117.098	176.734	1.042.151	1.844.883
Poc freqüent	51.485	13.889	21.457	77.436	164.267
Total	3.709.974	899.307	1.310.407	7.643.117	13.562.805

Taula 7.7: Mode de transport segons motiu del viatge RMB

Motiu	Cotxe	Autobús	Metro	Altres	Total
Freqüent	27,55%	6,95%	9,99%	55,51%	100%
Casa	26,79%	6,16%	9,04%	58,01%	100%
Feina	27,58%	6,35%	9,58%	56,49%	100%
Poc freqüent	31,34%	8,46%	13,06%	47,14%	100%

Taula 7.8: Mode de transport segons motiu del viatge RMB (%)

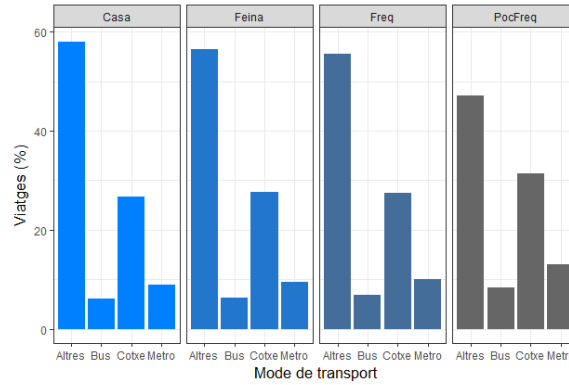


Figura 7.5: Mode de transport segons motiu del viatge RMB

Com s'ha dit abans les dades que fan referència als motius de viatge són difícilment comparables amb l'EMEF.

En els viatges segons el lloc de residència i mode de transport s'identifica el següent (Taula 7.9, Taula 7.10 i Figura 7.6):

- Els modes de transport privat i no motoritzat són els que utilitzen més els viatgers: un 70% els residents a Barcelona, un 83% els de la Primera Corona i un 96% els de la Regió Metropolitana de Barcelona del total de viatges diaris.
- Els residents a Barcelona són els que utilitzen més el transport públic en un 30% aproximadament dels viatges que fan durant el dia.

Residència	Cotxe	Autobús	Metro	Altres	Total
BCN	1.130.185	651.826	765.981	2.242.176	4.790.168
PC	1.304.690	178.558	401.529	1.622.287	3.507.064
RMB	1.275.099	68.923	142.897	3.778.654	5.265.573
Total	3.709.974	899.307	1.310.407	7.643.117	13.562.805

Taula 7.9: Mode de transport segons lloc de residència RMB

Residència	Cotxe	Autobús	Metro	Altres	Total
BCN	23,59%	13,61%	15,99%	46,81%	100%
PC	37,20%	5,09%	11,45%	46,26%	100%
RMB	24,22%	1,31%	2,71%	71,76%	100%

Taula 7.10: Mode de transport segons lloc de residència RMB (%)

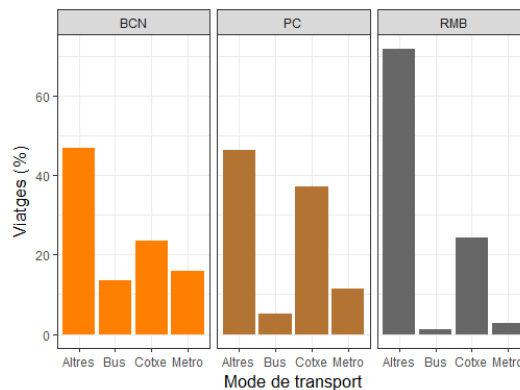


Figura 7.6: Mode de transport segons lloc de residència RMB

En aquest cas, el que no es pot comparar és el que fa referència als residents a la RMB que no inclou PC, perquè l'EMEF no treballa amb aquest àmbit. Per residents a Barcelona i PC (no inclou Barcelona), els percentatges modals són bastant semblants a l'EMEF.

Un cop analitzat els viatges a la Regió Metropolitana de Barcelona, s'ha observat que la quantitat de viatges realitzats dins la Primera Corona és un 61% del total, per tant ara s'analitzaran els fluxos dins de la PC.

7.3. Dades de mobilitat a la Primera Corona

En la Primera Corona és més difícil comparar els resultats amb l'EMEF, perquè l'enquesta no es centra tant en aquest àmbit, sinó amb l'Àrea Metropolitana de Barcelona.

Si es vol veure la matriu origen-destinació dels municipis de la Primera Corona mireu a l'Apèndix (A), allà es detallen els viatges entre cadascun dels municipis.

A la Primera Corona es realitzen de mitjana uns 7,9 milions de viatges (Taula 7.11), on un 54% es fan amb origen i destinació a la ciutat de Barcelona, un 27% a la PC (no inclou Barcelona) i un 18% dels viatges es realitzen entre aquestes dues zones (Figura 7.7).

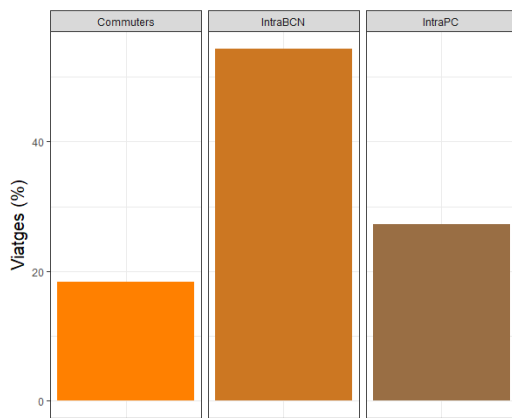


Figura 7.7: Àrea de mobilitat PC

Zona	Viatges	Percentatge
IntraBCN	4.288.894	54,30%
IntraPC	2.155.116	27,28%
Commuters	1.454.965	18,42%
Total	7.898.975	100%

Taula 7.11: Àrea de mobilitat PC

En aquests resultats passa més o menys el mateix que en els primers de la RMB, que en l'EMEF hi ha més viatges, 2 milions aproximadament, els factors que han pogut fer augmenta tants de desplaçament, són els mateixos que abans: s'apliquen factors de correcció i que els no residents a la RMB també estan comptabilitzats.

El següent pas es desagreguen les dades anteriors segons el motiu del viatge. I s'ha pogut observar el següent (Taula 7.12):

- Els viatges realitzats dins de Barcelona tenen un poc més del 50% del total de cada motiu, seguidament dels viatges dintre de la Primera Corona (no inclou Barcelona) amb un 27%, aproximadament, en cada motiu, i finalment els viatges de connexió (*commuters*) tenen sobre un 18%, en cadascun d'ells (Taula 7.13 i Figura 7.8).

Motiu	IntraBCN	IntraPC	Commuters	Total
Freqüent	2.421.301	1.096.215	767.543	4.285.059
Casa	1.245.046	749.312	465.693	2.460.051
Feina	561.683	287.229	200.039	1.048.950
Poc Freqüent	60.864	22.361	21.690	104.915
Total	4.288.894	2.155.117	1.454.965	7.898.976

Taula 7.12: Àrea de mobilitat segons el motiu del viatge PC

Motiu	IntraBCN	IntraPC	Commuters	Total
Freqüent	56,51%	25,58%	17,91%	100%
Casa	50,61%	30,46%	18,93%	100%
Feina	53,55%	27,38%	19,07%	100%
Poc Freqüent	58,01%	21,31%	20,67%	100%

Taula 7.13: Àrea de mobilitat segons el motiu del viatge PC (%)

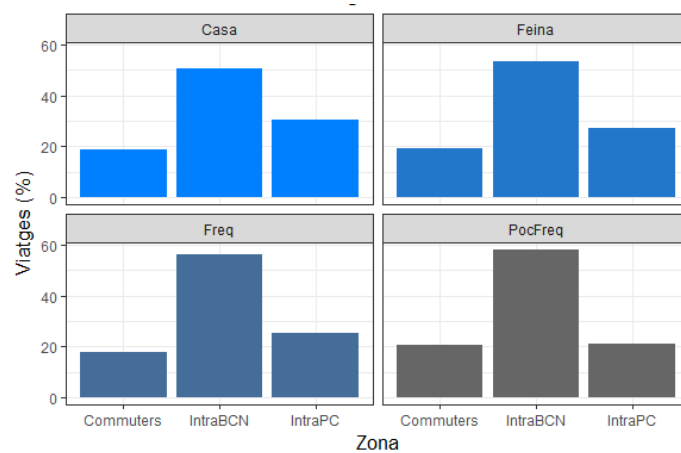


Figura 7.8: Àrea de mobilitat segons el motiu del viatge PC

Tal com s'ha realitzat per la RMB, ara es fa per la PC segons el lloc de residència, i el que s'observa és:

- Quasi tots els que són residents de Barcelona realitzen els viatges dins la ciutat de Barcelona amb una incidència del 84% (Taula 7.15).
- La majoria dels residents a la PC (no inclou Barcelona) viatgen dins d'aquesta zona amb un 62% dels viatges diaris, també hi ha un 27% dels viatges que són de connexió entre la PC i Barcelona (Figura 7.9).
- Els residents a la RMB (que no pertanyen a la PC) vinculat amb desplaçament a la Primera Corona, faciliten un volum molt minoritari de viatges a la PC, però el 58% dels viatges d'aquest grup que es realitzen són interns a la ciutat de Barcelona (Taula 7.14 i Figura 7.9).

Residència	IntraBCN	IntraPC	Commuters	Total
BCN	3.681.887	151.581	546.671	4.380.139
PC	352.893	1.899.404	831.600	3.083.897
RMB	254.115	104.132	76.694	434.941
Total	4.288.894	2.155.117	1.454.965	7.898.977

Taula 7.14: Àrea de mobilitat segons el lloc de residència PC

Residència	IntraBCN	IntraPC	Commuters	Total
BCN	84,06%	3,46%	12,48%	100%
PC	11,44%	61,59%	26,97%	100%
RMB	58,43%	23,94%	17,63%	100%

Taula 7.15: Àrea de mobilitat segons el lloc de residència PC (%)

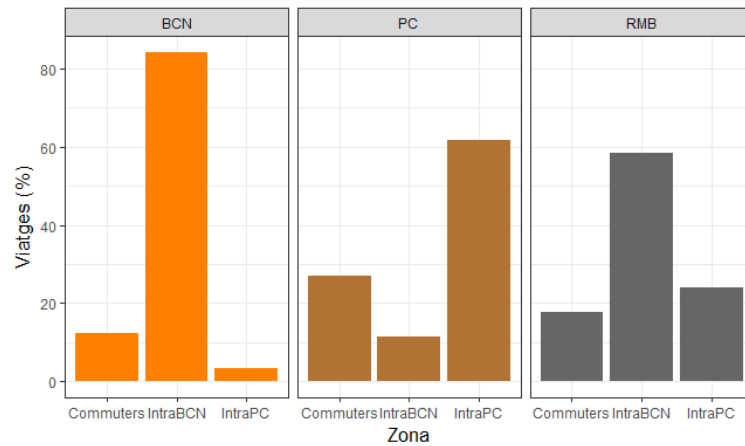


Figura 7.9: Àrea de mobilitat segons el lloc de residència PC

Es mira quin mode de transport s'utilitza en els viatges en la PC, i el més rellevant és:

- Dintre de Barcelona els viatges es realitzen a parts iguals entre el transport motoritzat (cotxe, autobús i metro) i el no motoritzat (altres) (Taula 7.16).
- El 79,76% dels viatges que connecten la Primera Corona i Barcelona utilitzen transport motoritzat, on cal destacar l'ús del transport privat (cotxe), ja que té un 42,89% dels viatges totals en aquest flux (Taula 7.17).
- A la Primera Corona (no inclou Barcelona), el 60,2% dels desplaçaments són amb transport no motoritzat (Figura 7.10).

Zona	Viatges	Cotxe	Autobús	Metro	Altres
IntraBCN	4.288.897	747.753	667.481	685.961	2.187.702
IntraPC	2.155.116	723.791	44.863	89.091	1.297.371
Commuters	1.454.963	623.974	150.857	385.719	294.412
Total	7.898.976	2.095.518	863.201	1.160.771	3.779.485

Taula 7.16: Àrea de mobilitat segons el mode de transport PC

Zona	Cotxe	Autobús	Metro	Altres	Total
IntraBCN	17,43%	15,56%	15,99%	51,01%	100%
IntraPC	33,58%	2,08%	4,13%	60,20%	100%
Commuters	42,89%	10,37%	26,51%	20,24%	100%

Taula 7.17: Àrea de mobilitat segons el mode de transport PC (%)

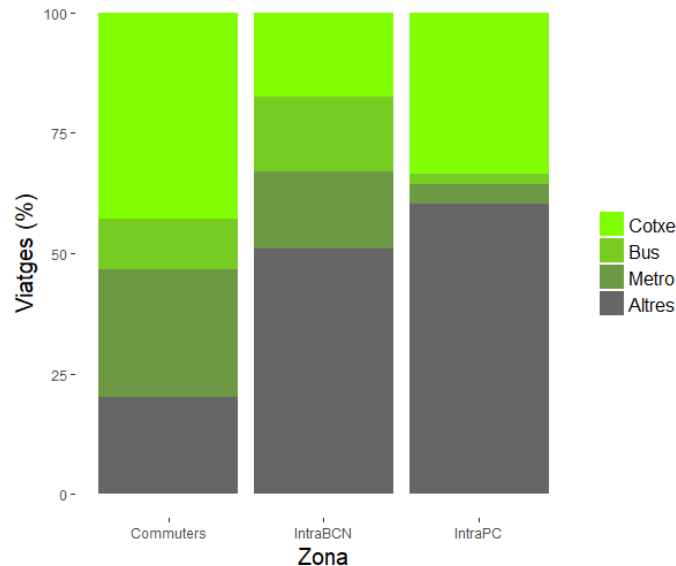


Figura 7.10: Àrea de mobilitat segons el mode de transport PC

En aquests resultats, observant també els de l'EMEF, el que és més remarcable és a la PC que no inclou Barcelona, ja que la diferència de l'ús del transport públic entre els dos resultats és molt gran: amb les dades de Kineo hi ha 133.954 viatges i en els resultats de l'EMEF hi ha 813.533 desplaçaments.

Es passa a analitzar els viatges de la PC segons el mode de transport i segons el motiu, on no hi ha massa cosa per destacar (Taula 7.18), ja que es comporten sense diferències remarcables respecte del mode o del motiu, respectivament, per tant hi ha independència entre el mode i el motiu dins dels desplaçaments a la Primera Corona, però el que més es pot destacar és:

- En els viatges per motiu poc freqüent, dintre de Barcelona i amb cotxe, és el grup on augmenta més respecte als altres casos, on hi ha un 11% dels viatges que es fan per aquest motiu (Taula 7.19 i Figura 7.11).

Motiu	Zones	Viatges	Cotxe	Autobús	Metro	Altres	Total
Freqüent	IntraBCN	2.421.303	420.210	373.728	389.607	1.237.758	4.285.059
	IntraPC	1.096.215	374.458	23.925	45.534	652.297	
	Commuters	767.541	331.450	79.979	201.290	154.823	
Casa	IntraBCN	1.245.047	218.118	198.649	196.613	631.667	2.460.052
	IntraPC	749.312	245.682	14.506	30.980	458.143	
	Commuters	465.693	196.274	47.609	124.472	97.339	
Feina	IntraBCN	561.683	97.878	84.755	88.111	290.939	1.048.950
	IntraPC	287.229	94.764	5.812	11.401	175.252	
	Commuters	200.038	87.101	20.999	53.958	37.980	
Poc Freqüent	IntraBCN	60.864	11.547	10.349	11.630	27.339	104.916
	IntraPC	22.361	8.887	619	1.176	11.678	
	Commuters	21.690	9.149	2.271	6.000	4.271	
Total		7.898.976	2.095.518	863.201	1.160.772	3.779.486	7.898.977

Taula 7.18: Mode de transport segons motiu PC

Motiu	Zona	Cotxe	Autobús	Metro	Altres	Total
Freqüent	IntraBCN	9,81%	8,72%	9,09%	28,88%	100%
	IntraPC	8,74%	0,56%	1,06%	15,22%	
	Commuters	7,74%	1,87%	4,70%	3,61%	
Casa	IntraBCN	8,87%	8,08%	7,99%	25,67%	100%
	IntraPC	9,99%	0,59%	1,26%	18,61%	
	Commuters	7,98%	1,94%	5,06%	3,96%	
Feina	IntraBCN	9,33%	8,08%	8,40%	27,74%	100%
	IntraPC	9,03%	0,55%	1,09%	16,71%	
	Commuters	8,30%	2,01%	5,14%	3,62%	
Poc Freqüent	IntraBCN	11,01%	9,86%	11,09%	26,06%	100%
	IntraPC	8,47%	0,59%	1,12%	11,13%	
	Commuters	8,72%	2,16%	5,72%	4,07%	

Taula 7.19: Mode de transport segons motiu PC (%)

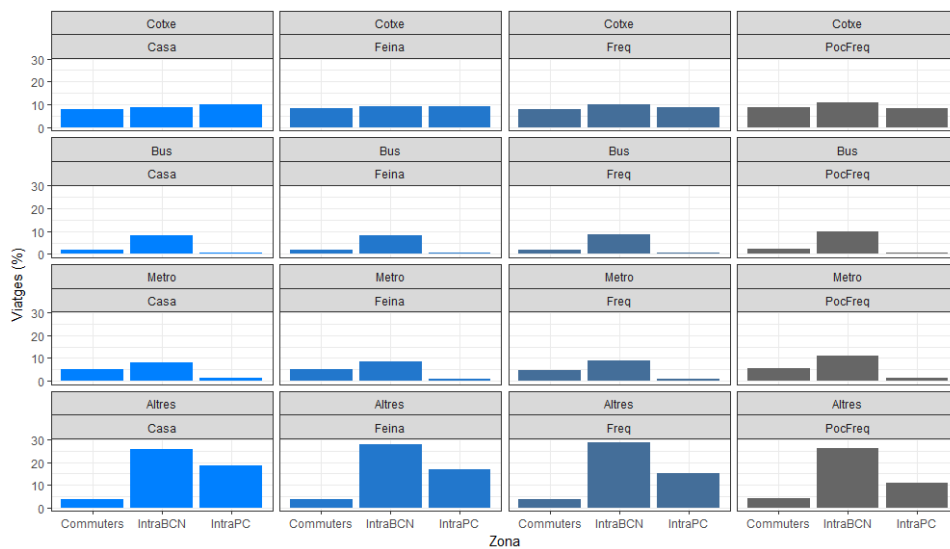


Figura 7.11: Mode de transport segons motiu PC

Per remarcar la importància dels modes motoritzats, a continuació es fa l'anàlisi dels desplaçaments per mode/motiu després d'eliminar el mode de transport no motoritzat (grup altres), doncs distorsiona bastant els resultats, en tractar-se de viatges curts i amb un volum al voltant del 50% del total.

El més interessant que s'ha observat és (Taula 7.20):

- Dintre de Barcelona viatgen un 64,4% en transport públic.
- En la Primera Corona (no inclou Barcelona) s'utilitza el transport privat en un 84,4% dels viatges (Taula 7.21).
- Els viatges que connecten aquests dos àmbits utilitzen el cotxe en un 53,8% i en un 33,3% el metro (Figura 7.12).

Zona	Viatges	Cotxe	Autobús	Metro
IntraBCN	2.101.195	747.753	667.481	685.961
IntraPC	857.745	723.791	44.863	89.091
Commuters	1.160.550	623.974	150.857	385.719
Total	4.119.490	2.095.518	863.201	1.160.771

Taula 7.20: Àrea de mobilitat transport motoritzat PC

Zona	Cotxe	Autobús	Metro	Total
IntraBCN	35,59%	31,77%	32,65%	100%
IntraPC	84,38%	5,23%	10,39%	100%
Commuters	53,77%	13,00%	33,24%	100%

Taula 7.21: Àrea de mobilitat transport motoritzat PC (%)



Figura 7.12: Àrea de mobilitat transport motoritzat PC

Les dades anteriors es desagreguen segons el motiu de viatge (Taula 7.22), i s'observa el següent:

- Els viatges per motiu tornar a casa, dintre de la Primera Corona (no inclou Barcelona), en cotxe representen un 19,3%, del total motoritzat, és en l'únic motiu que la PC supera a la ciutat de Barcelona en mode de transport motoritzat (Taula 7.23).

- També cal destacar que tots els viatges de connexió per qualsevol motiu amb transport públic són superiors als de dintre de la Primera Corona (Figura 7.13).

Motiu	Zona	Desplaçaments	Cotxe	Bus	Metro	Total
Freqüent	IntraBCN	1.183.545	420.210	373.728	389.607	2.240.181
	IntraPC	443.917	374.458	23.925	45.534	
	Commuters	612.718	331.450	79.979	201.290	
Casa	IntraBCN	613.380	218.118	198.649	196.613	1.272.903
	IntraPC	291.169	245.682	14.506	30.980	
	Commuters	368.354	196.274	47.609	124.472	
Feina	IntraBCN	270.744	97.878	84.755	88.111	544.779
	IntraPC	111.977	94.764	5.812	11.401	
	Commuters	162.058	87.101	20.999	53.958	
Poc Freqüent	IntraBCN	33.526	11.547	10.349	11.630	61.627
	IntraPC	10.682	8.887	617	1.175	
	Commuters	17.420	9.149	2.271	6.000	
Total		4.119.490	2.095.518	863.201	1.160.771	4.119.490

Taula 7.22: Transport motoritzat segons el motiu del viatge PC

Motiu	Zona	Cotxe	Autobús	Metro	Total
Freqüent	IntraBCN	18,76%	16,68%	17,38%	100%
	IntraPC	16,72%	1,07%	2,03%	
	Commuters	14,80%	3,57%	8,99%	
Casa	IntraBCN	17,14%	15,61%	15,44%	100%
	IntraPC	19,30%	1,14%	2,43%	
	Commuters	15,42%	3,74%	9,78%	
Feina	IntraBCN	17,97%	15,56%	16,17%	100%
	IntraPC	17,39%	1,08%	2,09%	
	Commuters	15,99%	3,85%	9,90%	
Poc Freqüent	IntraBCN	18,74%	16,79%	18,87%	100%
	IntraPC	14,42%	1,00%	1,91%	
	Commuters	14,85%	3,68%	9,74%	

Taula 7.23: Transport motoritzat segons el motiu del viatge PC (%)

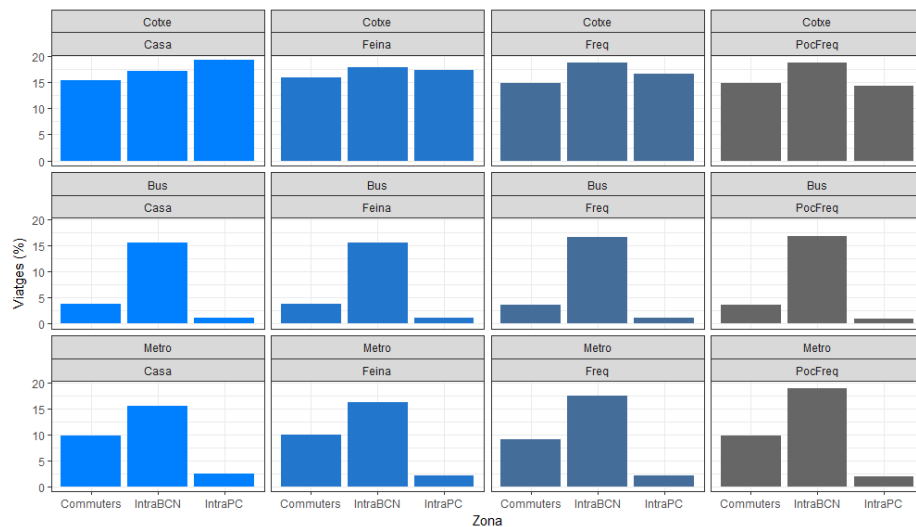


Figura 7.13: Àrea de mobilitat transport motoritzat PC

Un indicador interessant d'analitzar és si s'utilitza més el transport públic o privat, per això s'ha calculat una ràtio de transport públic/privat. Els resultats són (Taula 7.24):

- Els viatges dintre de Barcelona són els que més es realitzen en transport públic, amb una ràtio de 1,81, o sigui, quasi s'utilitza el doble el transport públic que el privat.
- La majoria dels viatges dintre la Primera Corona (no inclou Barcelona) es fan en transport privat, amb una ràtio de 0,19.
- Els viatges amb connexió entre aquestes dues zones són superiors en transport privat, amb una ràtio de 0,86.
- La ràtio públic/privat del total de viatges dins la Primera Corona és de 0,97, això vol dir que hi ha lleugerament més viatges en transport privat que públic.

Zona	Públic	Privat	Total	Ràtio
IntraBCN	1.353.442	747.753	2.101.195	1,81
IntraPC	133.954	723.791	857.740	0,19
Commuters	536.576	623.974	1.160.550	0,86
Total	2.023.972	2.095.518	4.119.490	0,97

Taula 7.24: Ràtio transport públic/privat

En aquest punt hi ha aspectes interessants per comparar amb els resultats de l'EMEF, les ràtios públic/privat d'ús modal en els viatges dins de Barcelona són bastants semblants, en canvi els viatges de connexió i els de dins de la PC no inclou Barcelona són bastant diferents, ja que a les dades usades pel VML la ràtio és de 0,86 i 0,19, respectivament, i a l'EMEF és de 1,25 i de 0,48 respectivament. O sigui, la diferència és bastant gran, sobretot als viatges de connexió que a un lloc surt que s'utilitza més el transport públic i a l'altre el privat. Aquest detall implica que s'ha de considerar amb cautela l'ús de la mobilitat modal inferida a partir de les dades de KINEO i fusionada amb dades EMIT, les matrius modals en transport privat hauran d'ajustar-se durant el procés de calibratge del VML usant dades de càrregues de les línies subministrades pels operadors. A Barcelona, el repartiment modal sembla prou acurat i requereix menys intervenció en el procés de calibratge.

Un cop analitzats els fluxos territorials per la Regió Metropolitana de Barcelona i per la Primera Corona segons mode de transport, motiu del viatge i lloc de residència, s'estudia la distribució horària dels viatges a la RMB en general, segons el motiu del viatge i el lloc de residència. Les dades de telefonia mòbil tenen una granularitat horària que es parteix de la hipòtesi que ha de ser de major qualitat que la que es podria obtenir a partir de l'expansió de les dades obtingudes a partir d'enquesta tradicional.

7.4. Distribució horària dels desplaçaments a la RMB

Observant les franges horàries (Figura 7.14) a gran escala es pot veure que es distribueixen en tres períodes, que són:

- De les 7 h a les 10 h, és quan la gent va a treballar, estudiar, entre altres motius.
- De les 13 h a les 15 h, és l'hora d'anar a dinar, canvi de torn a la feina o a la universitat, entre altres causes.
- De les 17 h a les 20 h, és quan la majoria de persones acaben les seves responsabilitats del dia i tornen cap a casa, com ara: els treballadors, estudiants, o els viatges amb motiu no ocupacional.

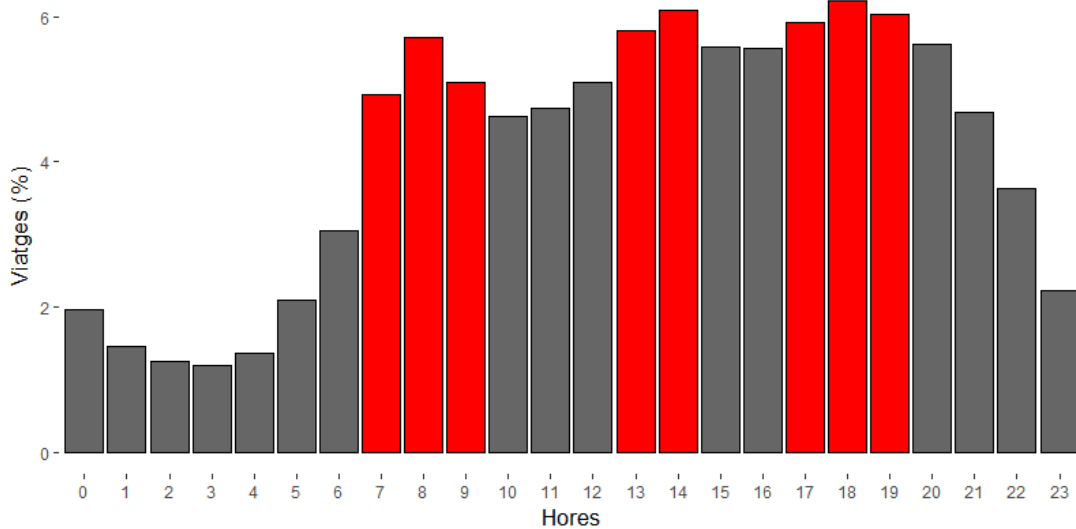


Figura 7.14: Franja horària RMB

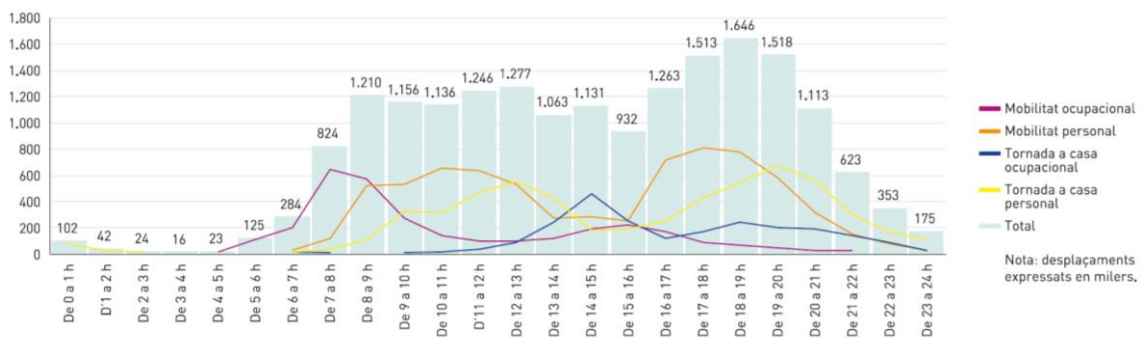


Figura 7.15: Franja horària general i segons motiu, segons l'EMEF.

Font: (Autoritat del Transport Metropolità 2015)

La gran diferència amb l'EMEF és que amb les dades de Kineo es remarquen molt bé tres franges, en canvi, les de l'EMEF (Figura 7.15) només té una punta de major intensitat de desplaçaments a la tarda, i al matí té una certa tendència a l'homogeneïtzació. Això és pel fet que les enquestes no solen capturar els viatges curts, en canvi a les dades procedents de Kineo es comptabilitzen tots.

En les franges horàries segons el motiu del viatge s'observa (Figura 7.16):

- La mobilitat per motius personals es separa en dos grups: llocs que es visiten freqüentment (Freq) i poc freqüentment (PocFreq). En el primer grup la majoria de desplaçaments es produeixen al llarg del dia entre les 7 i les 19 hores, que correspon a un 80% dels viatges diaris. L'altre grup, els viatges són més o menys constants al llarg de tot el dia.
- La franja de major concentració de viatges per motius laborals (Feina) és entre les 7 i les 10 hores del matí (aproximadament el 30% de la mobilitat ocupacional diària).
- Les tornades a casa presenten una distribució amb dues franges de major intensitat, al migdia de 13 a 15 hores i a la tarda de les 18 a les 21 hores. On aquestes últimes representen un 43% de les tornades a casa en tot el dia.

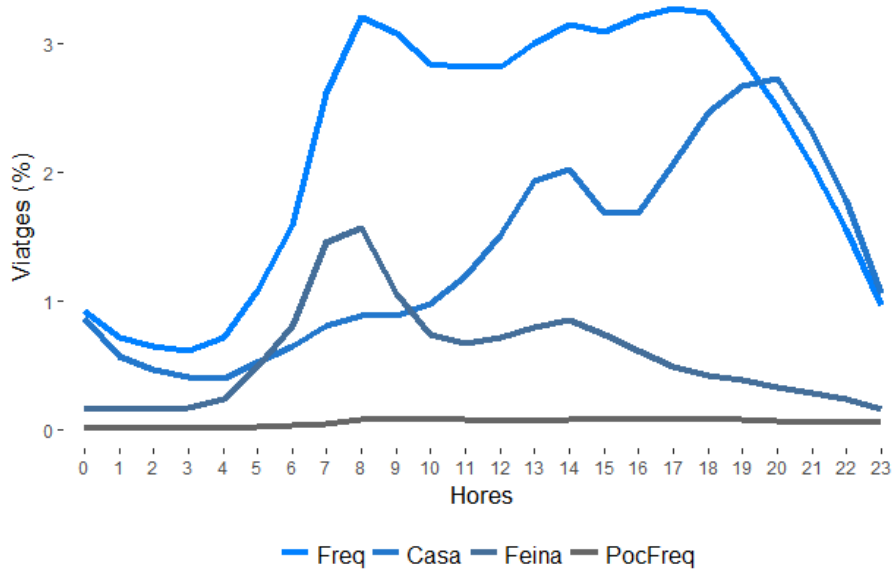


Figura 7.16: Franja horària segons el motiu del viatge RMB

En aquest cas, a partir de les franges horàries per motiu de viatge sí que hi ha una certa similitud amb l'EMEF, el motiu anar a la feina té una intensitat a primera hora (als dos casos), les tornades a casa hi ha una intensitat al migdia i una altra més forta a la tarda-nit. I la gran diferència és que el motiu freqüent mostra una certa tendència a l'homogeneïtat durant tot el dia (dades Kineo), i per l'altra part només té una major intensitat a primera hora del matí (EMEF).

I segons el lloc de residència (Figura 7.17):

- Es comporta més o menys igual per a les tres zones (Barcelona, Primera Corona i Regió Metropolitana de Barcelona). Hi ha tres períodes amb una major intensitat de desplaçaments: matí, migdia i tarda (tal com s'han vist a escala global).

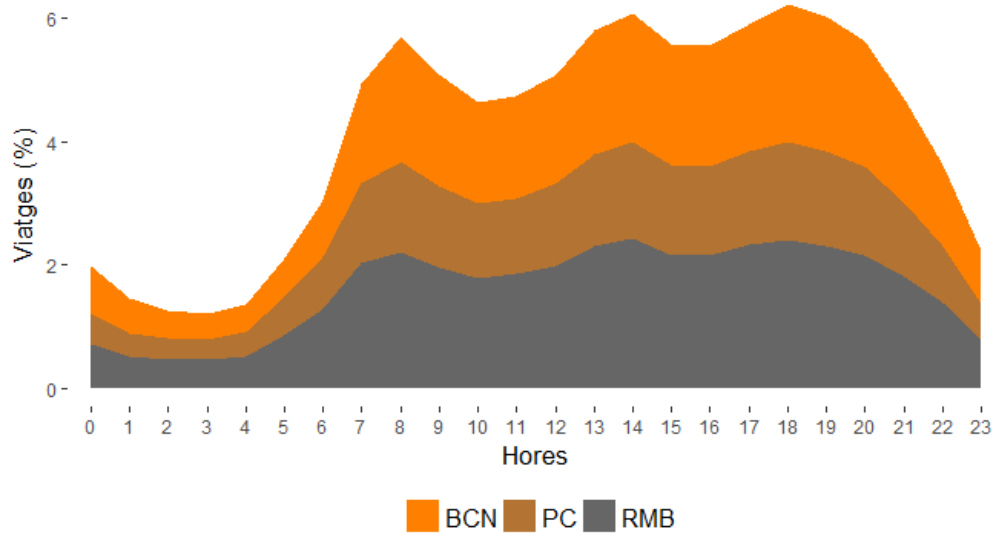


Figura 7.17: Franja horària segons el lloc de residència RMB

7.5. Mapa de calor

Un mapa de calor és una representació gràfica qualitativa de les dades, on els valors dels individus continguts en una matriu es representen com colors.

En aquest cas, s'adreça a veure en un mapa de calor la quantitat de viatges entre municipis, on el color vermell indica que hi ha una quantitat més elevada, tendint al verd en descendir els totals OD, segons una escala de colors logarítmica respecte als viatges OD i el gris que hi ha absència de viatges. Aquests mapes es representen per la Primera Corona, entre municipis origen i destinació, on les cel·les representen la quantitat de viatges realitzats en un dia, primer es considera els viatges totals i després per a cadascun dels modes de transport.

En el primer cas (Figura 7.18), que són els viatges totals, s'observa que on hi ha més moviment és a Barcelona, més concretament als viatges amb origen i destinació Barcelona. També es pot veure que on hi ha menys viatges és entre els municipis més allunyats entre ells, per exemple origen Tiana, destinació Sant Feliu de Llobregat.

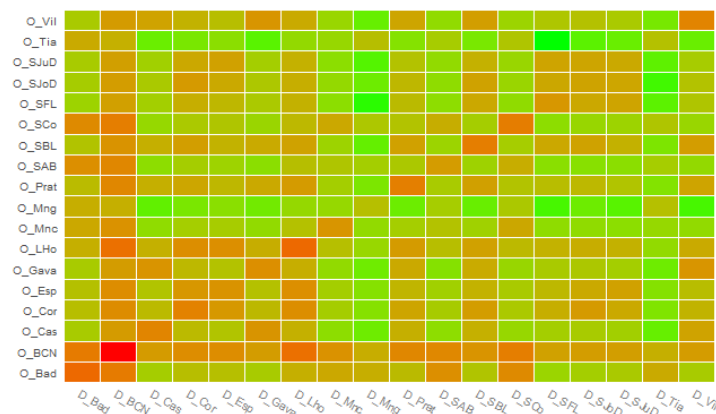


Figura 7.18: Mapa de calor OD municipis

Quan es fa representació per mode de transport, concretament el cotxe (Figura 7.19), s'observa que hi ha cel·les sense informació (color gris), aquestes cel·les corresponen majoritàriament a desplaçaments interiors d'alguns municipis, on origen i destinació és el mateix, o sigui que no es comptabilitzen desplaçaments en transport privat dins d'aquests municipis.

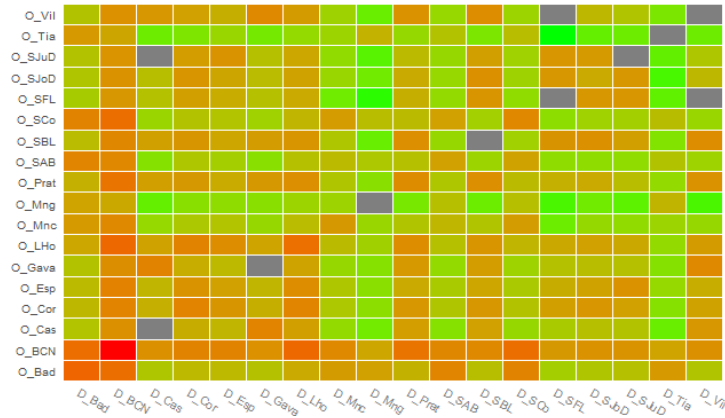


Figura 7.19: Mapa de calor OD Municipis - Cotxe

El transport públic té moltes cel·les sense informació (Figura 7.20 i Figura 7.21), moltes d'aquestes corresponen a les dues figures. També algunes d'elles corresponen, sobretot al metro, a municipis que no tenen aquest medi per desplaçar-se dins del mateix municipi.

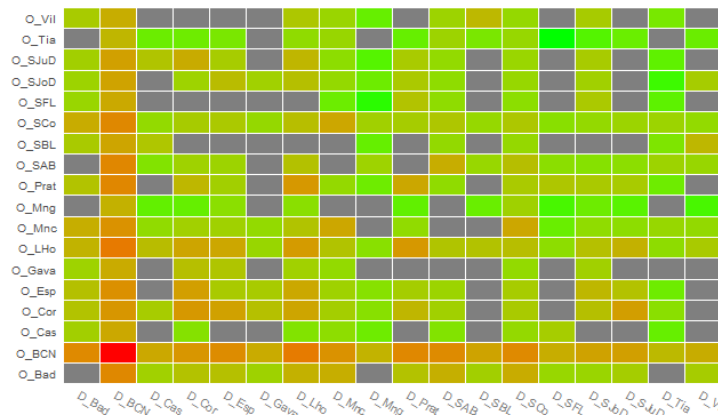


Figura 7.20: Mapa de calor OD municipis – Autobús

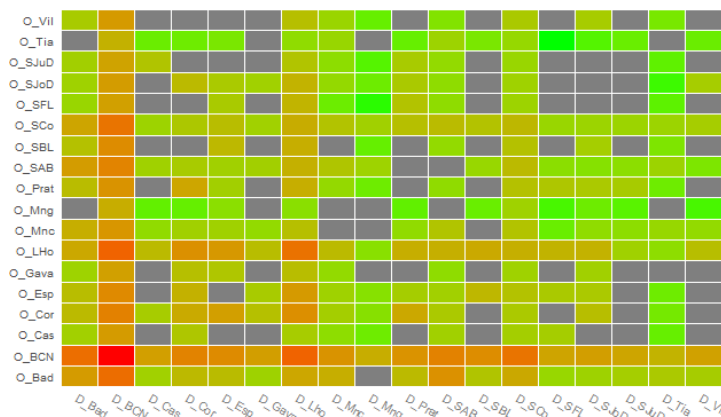


Figura 7.21: Mapa de calor OD municipis - Metro

relaciona amb els valors propis de la matriu de correlacions entre les variables actives (Components Principals Normalitzades), en aquest cas només n'hi ha un superior a 1, que és el que fa referència al primer eix principal. Però, que un eix capturi tanta informació i els altres tant poca pot ser a causa que hi hagi relacions OD atípiques (outliers), que són valors bastant diferents de la majoria per culpa del gran volum de viatges i que faciliten correlacions molt properes a 1 entre pràcticament tots els parells d'hores. La finalitat de l'anàlisi horària és validar si els perfils horaris de matí, hora vall, tarda i nit estan ben reflectits, molt correlacionats dins de cada franja i diferenciades les diferents franges, per això es descarten les relacions OD de gran volum que distorsionen aquesta lectura.

L'estudi de la selecció de les observacions atípiques es calcula mitjançant els percentils de les coordenades dels individus de l'ACP, calculat per a cadascun dels dos primers eixos principals. Es seleccionen aproximadament un 2,5% de les observacions (9336). Es comprova que aquestes parelles OD corresponent efectivament a les relacions OD indicades al punt anterior. Aquestes es passen al rol d'observacions suplementaries al realitzar el pròxim l'ACP. I la nova projecció és:

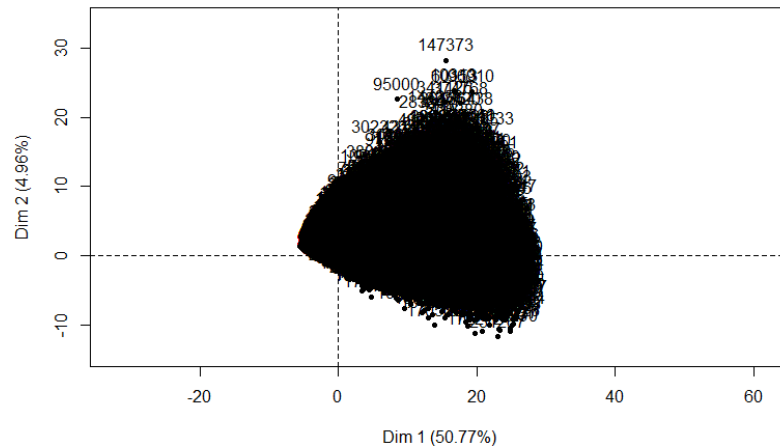


Figura 7.23: Projecció 1- 2 dels individus

Ara s'observa que quasi totes les observacions queden concentrades. Els dos primers eixos principals representen un 55% de la inèrcia total. En aquest cas, el criteri de Kaiser indica que hi ha tres eixos que tenen un valor propi superior a 1, per tant es realitza l'estudi en els quatre primers eixos principals. I la inèrcia acumulada dels 4 primers eixos principals és del 63%.

Seguidament es representen les variables numèriques actives segons la qualitat de la representació en els dos primers eixos principals:

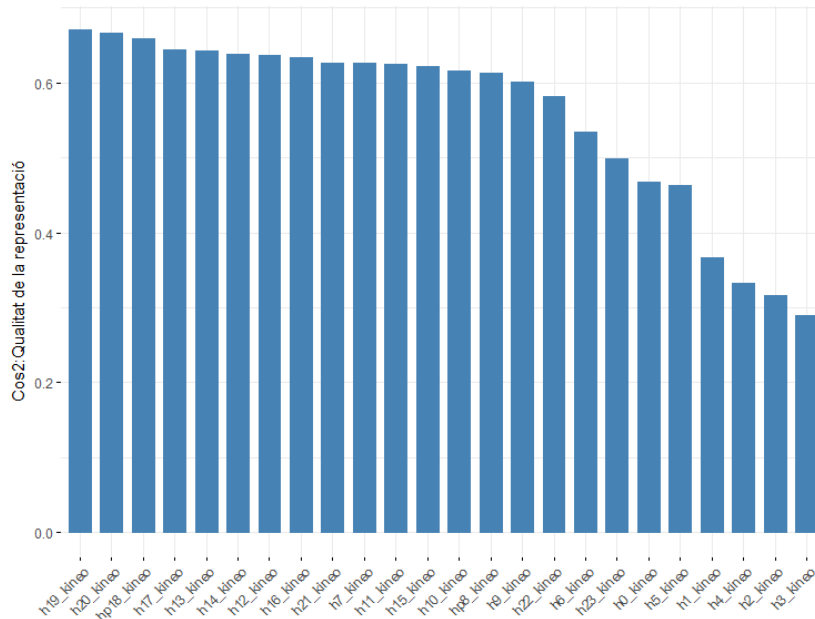


Figura 7.24: Cos^2 de les variables numèriques als eixos 1-2

Com s'observa les franges horàries que pertanyen a la matinada són les menys correlacionades amb els dos primers eixos.

Ara, es projecten les variables numèriques als dos primers eixos principals:

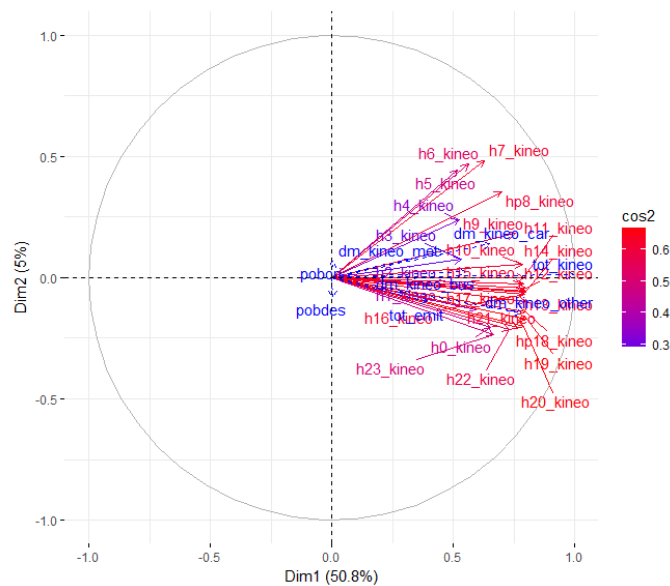


Figura 7.25: Projectió de les variables numèriques als eixos 1-2

S'observa que la variable latent del primer eix principal representa la quantitat de viatges realitzats. D'altra banda, la variable latent del segon eix podria estar relacionada amb la franja horària: hores correlatives tenen una alta correlació (angle petit), com ha de ser, hores en franja punta de matí es troben allunyades (en angle) de les hores punta de la tarda. Els resultats són satisfactoris de cara a la consistència indicada per les dades horàries segons els experts en el sector.

Ara es realitza el mateix procediment però per als eixos principals 3 i 4:

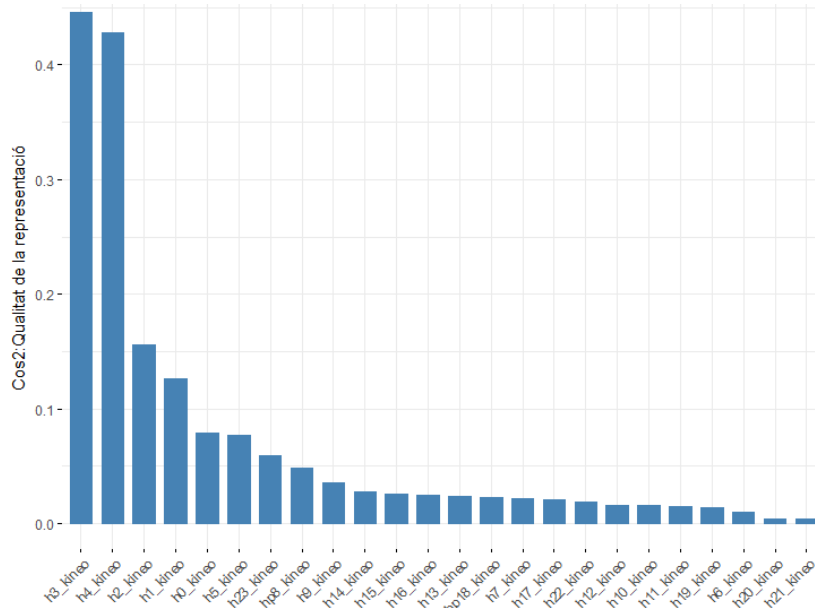


Figura 7.26: Cos^2 de les variables numèriques als eixos 3-4

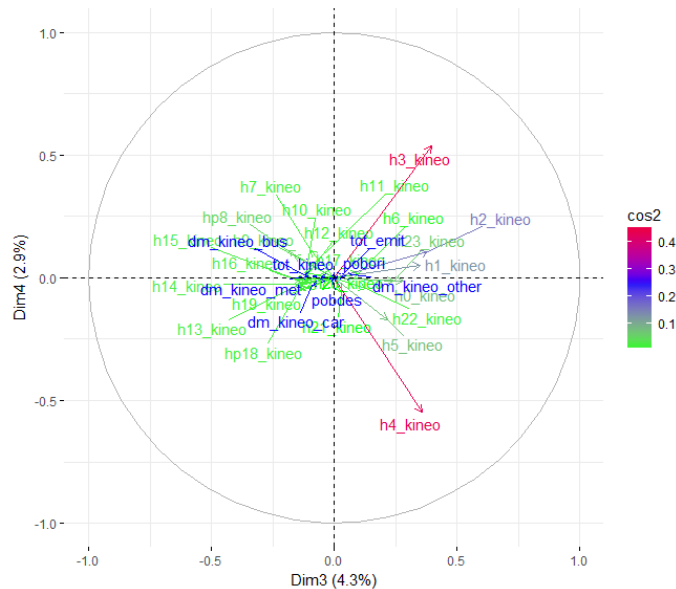


Figura 7.27: Projectió de les variables numèriques als eixos 3-4

En els eixos 3 i 4 les variables que tenen més informació (Figura 7.26) són les que fan referència a les franges horàries de 3 a 4 (h3_kineo) i de 4 a 5 (h4_kineo), clarament això representa el punt de ruptura entre la mobilitat nocturna per oci i l'inici de la mobilitat diürna de primera hora per mobilitat obligada. En la projecció de les variables numèriques en aquests eixos (Figura 7.27) s'observa que les variables h3_kineo i h4_kineo són ortogonals, per tant poc correlacionades. També la variable latent associada a l'eix 3 està feblement relacionada amb la modalitat, ja que oposa la motoritzada i no motoritzada segons indica la projecció de les variables suplementàries. En canvi, la projecció de l'eix 4 no s'associa en cap característica en concret.

Es projecten les variables categòriques al primer i segon eix principal per veure si ens poden donar més informació sobre les variables latents associades a aquests eixos. Però

les variables categòriques que s'ha trobat més representatives projectar són les que fan referència als municipis d'origen i de destinació, ja que la matriu de parelles OD fa referència a la Primera Corona. Les projeccions són:

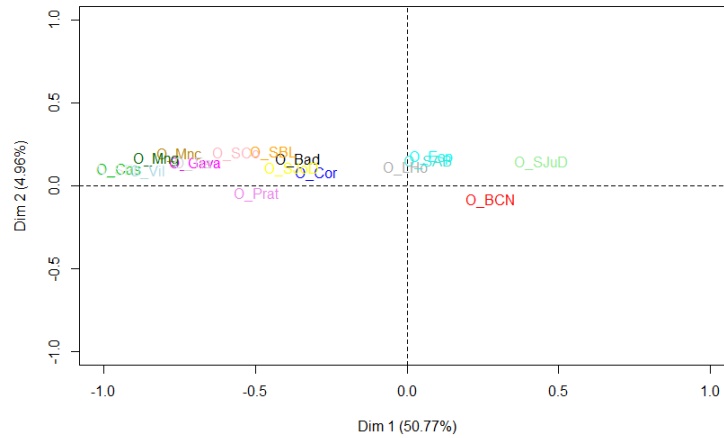


Figura 7.28: Projecció variable municipi origen eix 1-2

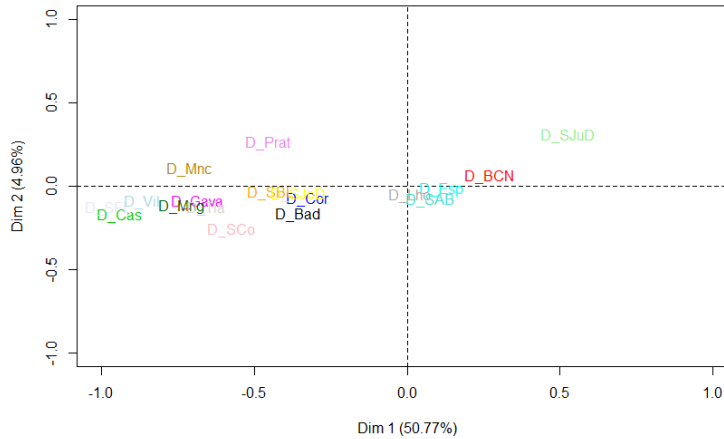


Figura 7.29: Projecció variable municipi destinació eix 1-2

En oposar el centre de gravetat de la generació de viatges de zones de Barcelona ciutat amb els centres de gravetat de les generacions dels altres municipis es desprèn un cert sentit del perfil direccional de la mobilitat: en augmentar la generació de viatges a Barcelona, es disminueixen les generacions de viatges als municipis de la Primera Corona com Cornellà, Badalona o Castelldefels. Una interpretació similar s'obté amb les atraccions.

7.7. Classificació a partir de l'ACP

La classificació mitjançant l'algoritme del K-means s'ha realitzat a partir dels resultats de l'ACP, amb els quatre primers eixos principals. El procediment a seguir és el que s'ha explicat a (5.2.1), on aquest algoritme s'ha aplicat mitjançant la funció per defecte que té R-Studio (*kmeans*). El nombre de classes que s'ha utilitzat és d'1 fins a 50 per a cadascuna de les execucions. Cadascuna de les execucions s'ha realitzat amb una assignació

aleatòria (llavor) diferent. Un cop realitzades, s'observa cadascuna de les variàncies intra-clústers per veure en quina classificació és la més petita i llavors es compara per a cada llavor. Finalment, la millor assignació aleatòria és 123 i el nombre de classes és 15. El nombre de clúster elegit podria variar, depenen el criteri de l'analitzador, en aquest cas el criteri és: com la diferència de la variància intra-clústers entre la iteració 15 i 16 no varia significativament, i a partir d'aquesta iteració les diferències eren petites entre elles, per tant el nombre de clústers elegit és 15.

Les classes resultants són:

Clúster	Observacions	Clúster	Observacions	Clúster	Observacions
Classe 1	213851	Classe 6	3720	Classe 11	2600
Classe 2	54399	Classe 7	29592	Classe 12	3864
Classe 3	2829	Classe 8	5248	Classe 13	934
Classe 4	7628	Classe 9	4924	Classe 14	1936
Classe 5	8638	Classe 10	4550	Classe 15	19883

Taula 7.25: Clúster ACP

Després de realitzar diferents gràfics i taules amb les agrupacions anteriors, no s'ha trobat cap patró que permeti caracteritzar clarament cadascuna de les classes. La manca de dades complementàries a les de mobilitat dificulta enormement la caracterització de les classes.

7.8. Anàlisi de Correspondències Simples (ACS)

Les correspondències simples permeten trobar les relacions entre orígens i destinacions de taules de contingència que representen viatges OD entre àmbits, per motius, modes, etc. La interpretació dels resultats és molt il·lustrativa a simple vista en alguns casos i resulta sorprenent la facilitat de caracterització dels perfils OD entre àmbits geogràfics que permet.

Per a la realització de l'ACS s'ha optat per comparar taules de contingència amb significat nombre de viatges segons diverses variables categòriques caracteritzades. Després de fer diverses relacions s'han escollit les que donen uns resultats més interessants.

Les observacions seleccionades són les mateixes que l'ACP original, o sigui, les que pertanyen a la Primera Corona, però al principi no s'exclouen les observacions de la diagonal. Les primeres variables a estudiar són les taules de viatges totals diaris entre els municipis de la Primera Corona. Un cop creada la taula de contingència amb la variable resposta total de viatges realitzats segons Kineo, s'ha comprovat que aquestes dues variables són dependents, i tot seguit s'ha realitzat l'ACS, amb els següents resultats:

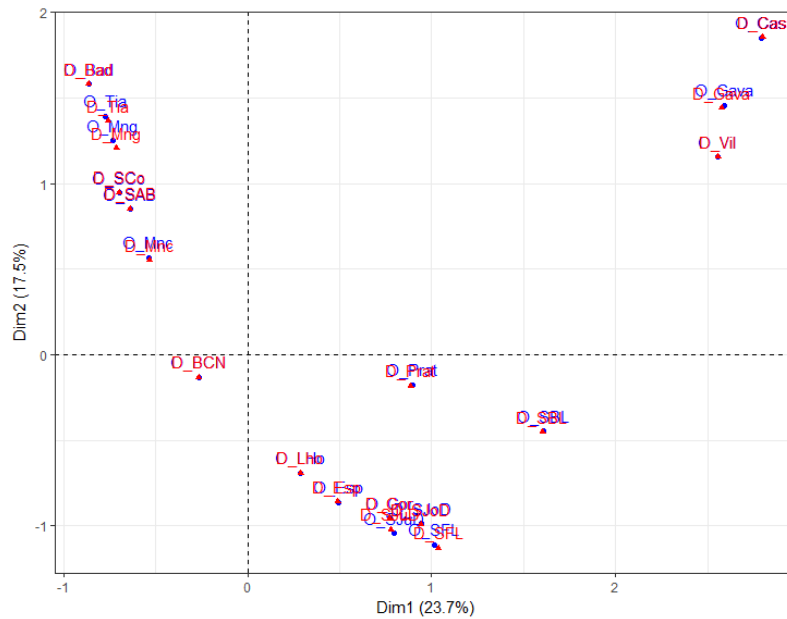


Figura 7.30: ACS municipis OD

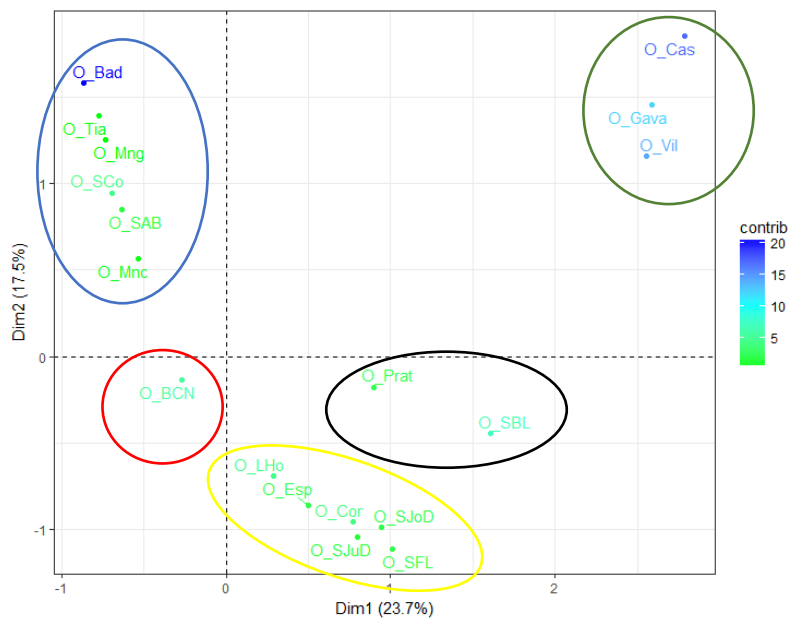


Figura 7.31: ACS municipis origen

Un cop projectades s'han realitzat dues gràfiques, ja que com es veu a la Figura 7.30 es superposen tant els orígens com les destinacions, i perquè més comprensible només es projecten les files de la taula de contingència (Figura 7.31), perquè tant els orígens com les destinacions tenen el mateix comportament en termes totals diaris. S'observa diferents grups on cadascun d'ells tenen el mateix patró, on Barcelona és el centre de gravetat (cercle vermell). I els grups són:

- els municipis que estan al nord de Barcelona (cercle blau).
- els municipis que estan al sud Barcelona (cercle groc).
- els que estan més allunyats de Barcelona (cercle verd).

- el Prat i Sant Boi de Llobregat (cercle negre) que es comporten diferent, podria ser perquè hi ha l'aeroport a prop.

Amb la mateixa taula de contingència que en el cas anterior, però es posen zeros a la diagonal, perquè es vol veure els viatges entre els municipis, o sigui, que l'origen i la destinació no és el mateix.

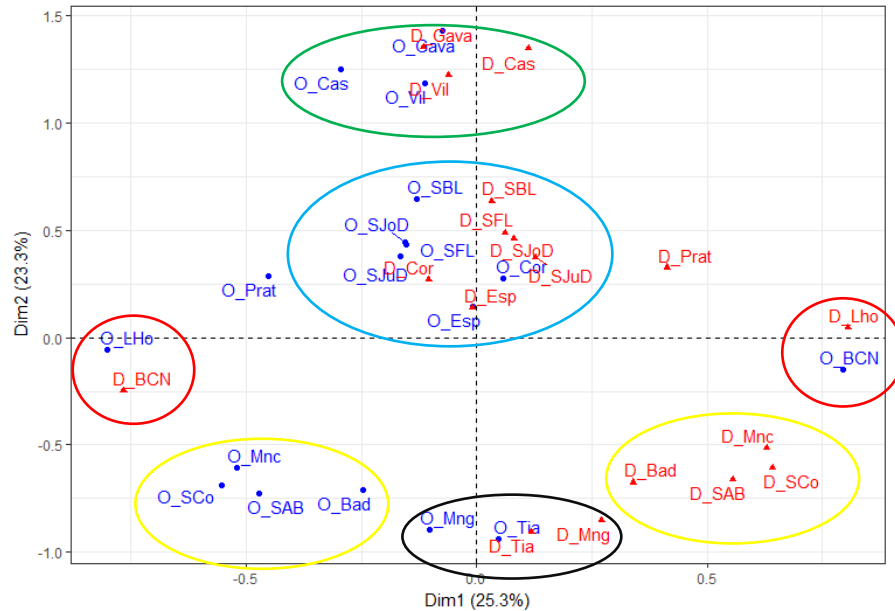


Figura 7.32: Viatges entre municipis OD

Ara s'observen patrons diferents (Figura 7.32), que són:

- Barcelona i l'Hospitalet de Llobregat (cercle vermell) tenen un patró oposat, o sigui, que quan un municipi és l'origen l'altre és la destinació.
- Montcada, Santa Coloma, Sant Adrià del Besòs i Badalona (cercle groc) tenen un perfil comú, però els orígens i destinacions d'aquests municipis estan oposats, però els orígens estan relacionats amb la destinació de Barcelona i viceversa.
- Montgat i Tiana (cercle negre) tenen un patró semblant. Concretament el municipi de Tiana, amb un intercanvi de viatges remarcable entre ells.
- Sant Boi, Sant Just Desvern, Cornellà, Esplugues i Sant Feliu (cercle blau) tenen un patró comú. Mirant amb més atenció el municipi d'Esplugues tant l'origen com la destinació estan relacionats, en canvi, els altres municipis es podria dir que els orígens de tots menys Cornellà tenen com a destinació Cornellà i viceversa.
- Castelldefels, Gavà i Viladecans (cercle verd) tenen un patró comú. Els municipis de Gavà i Viladecans tenen tant l'origen com la destinació el mateix municipi, en canvi, Castelldefels té un origen i una destinació més diversa.
- L'únic municipi que té un patró diferent de tots els altres és el Prat, les característiques específiques del qual es deuen a la presència de l'Aeroport.

Seguint amb les mateixes observacions anteriors, ara es vol veure si hi ha els mateixos patrons per a cadascuna de les franges del dia.

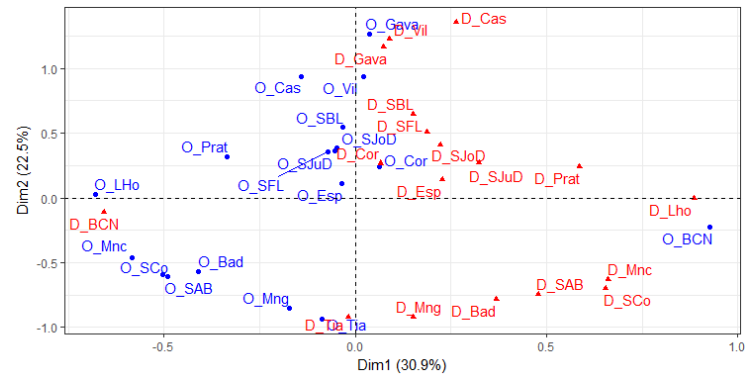


Figura 7.33: Viatges entre municipis, franja horària 7-10h

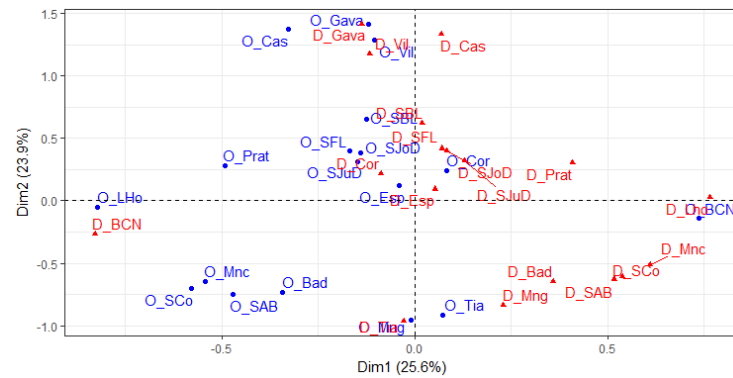


Figura 7.34: Viatges entre municipis, franja horària 13-15h

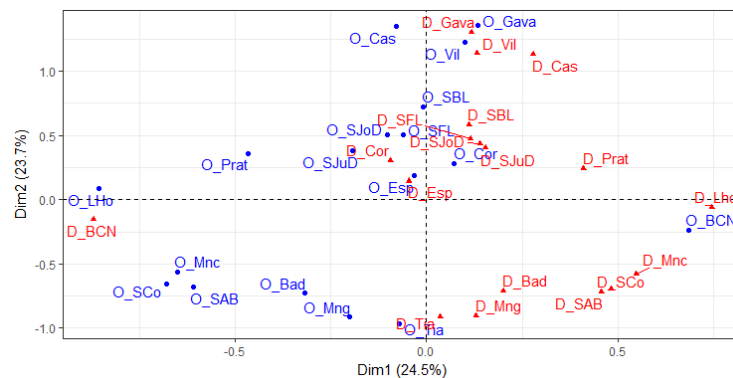


Figura 7.35: Viatges entre municipis, franja horària 17-20h

De la Figura 7.33 s'observa que la majoria d'orígens de municipis estan relacionats amb la destinació Barcelona, sobretot els que es situen al nord de Barcelona més L'Hospitalet, Tiana i Gavà en són una excepció, que tant l'origen com la destinació es donen en el mateix municipi. El segon eix enfronta els municipis del Barcelonès Nord amb els del Baix Llobregat.

A la franja horària que fa referència al migdia (Figura 7.34), alguns municipis del nord de Barcelona tenen el mateix patró que en l'anterior cas, però la diferència és que el municipi d'origen Montgat està molt correlacionat amb la destinació de Tiana. I els municipis del sud de Barcelona tenen un patró similar al de la Figura 7.32. I la franja horària de la tarda-nit (Figura 7.35) té un patró similar al de la Figura 7.32.

Un cop estudiats els municipis, també es va intentar estudiar les relacions de la mobilitat interior entre els barris de Barcelona, per tant només es seleccionen les parelles

OD que pertanyen a la ciutat de Barcelona. I es realitza el mateix procediment anterior. Només es posarà el gràfic per files, ja que tant els barris d'origen i destinació tenen un perfil comú en mobilitat total diària.

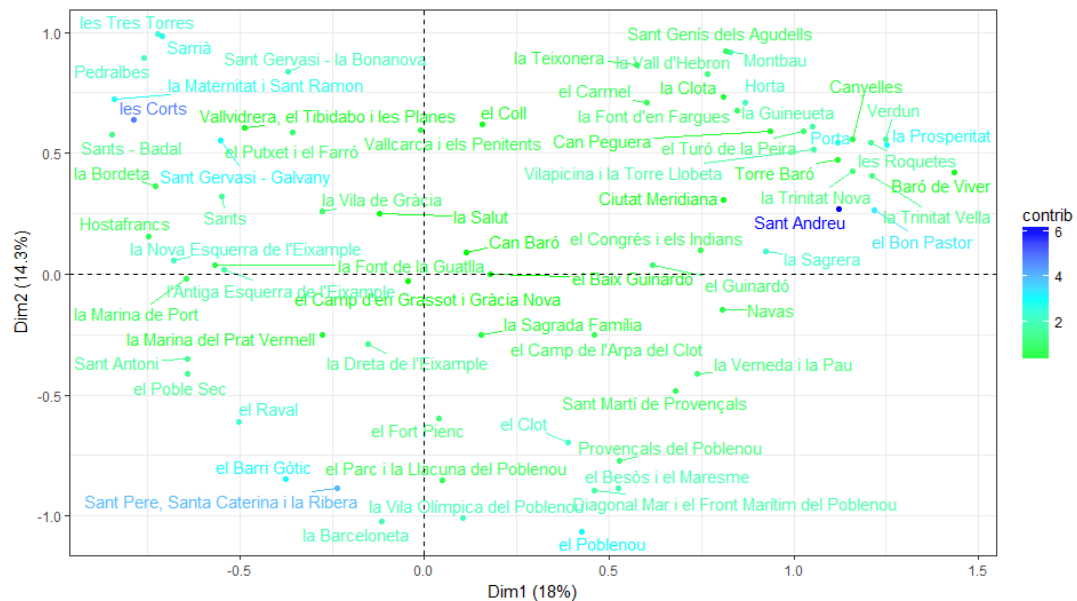


Figura 7.36: Viatges entre els barris de Barcelona

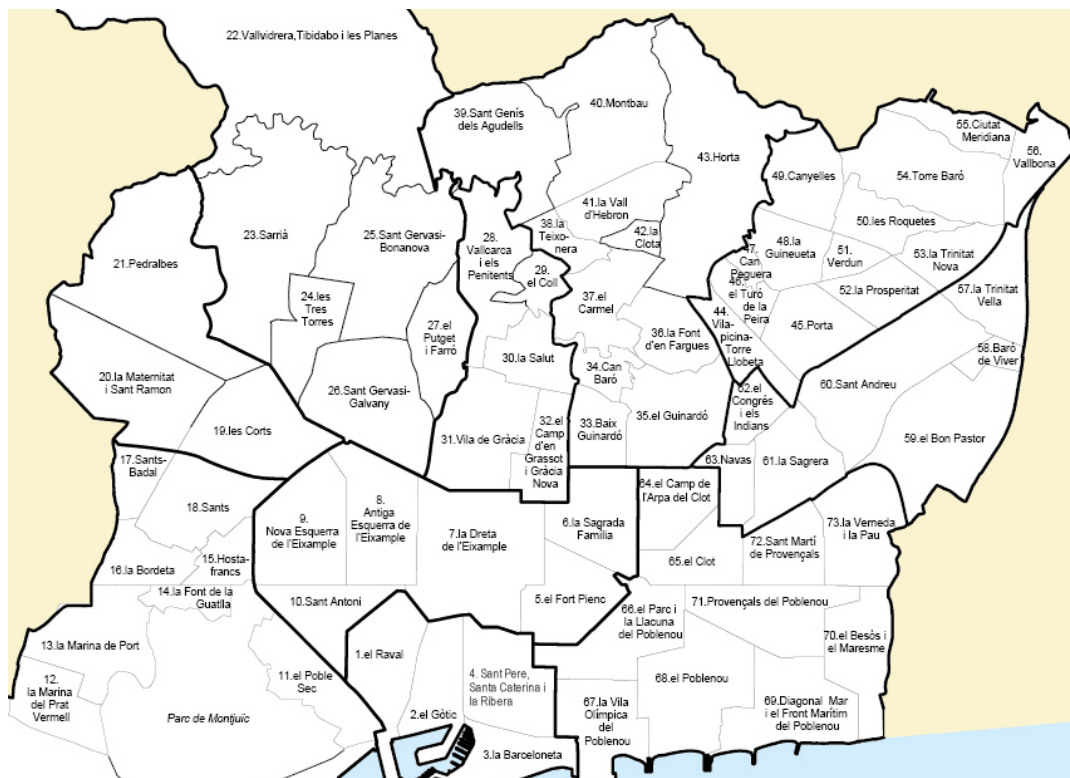


Figura 7.37: Mapa barris de Barcelona.

Font: <http://kezdonetblog.blogspot.com.es/2013/10/barris-de-barcelona.html>

La Figura 7.37 es posa per comparar la distribució real dels barris de Barcelona amb els resultats de l'ACS de la Figura 7.36, i es veu que hi ha molta similitud en la ubicació en la majoria dels barris, la qual cosa resulta molt reconfortant.

Es vol estudiar la ciutat de Barcelona depenent de quin municipi de la Primera Corona la visita. Es seleccionen les observacions d'origen de la Primera Corona menys de Barcelona i les observacions de destinació de la ciutat de Barcelona per districtes. I els resultats són els següents:

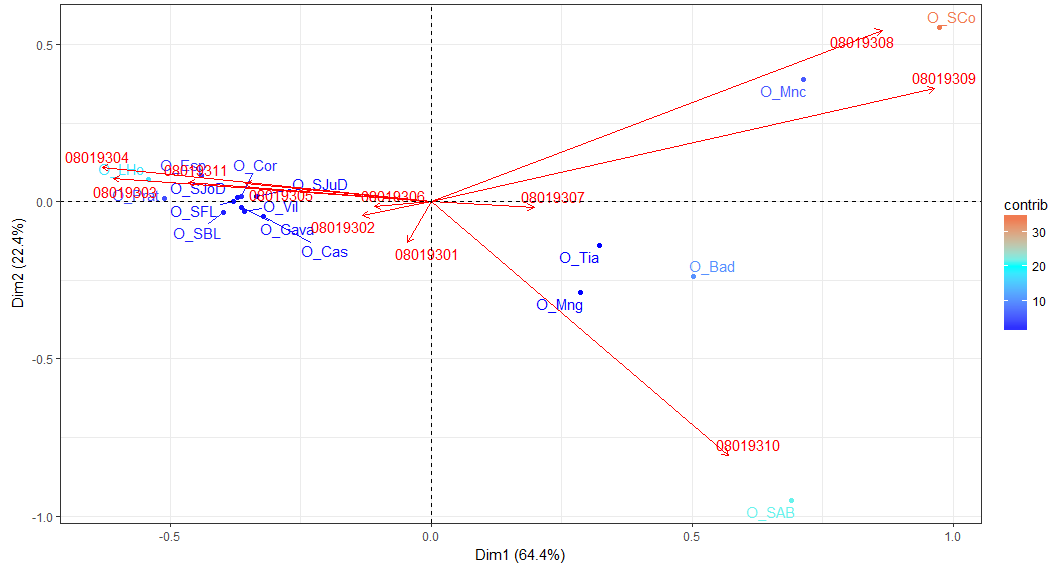


Figura 7.38: Viatges entre municipis origen de la PC i districtes destinació Barcelona

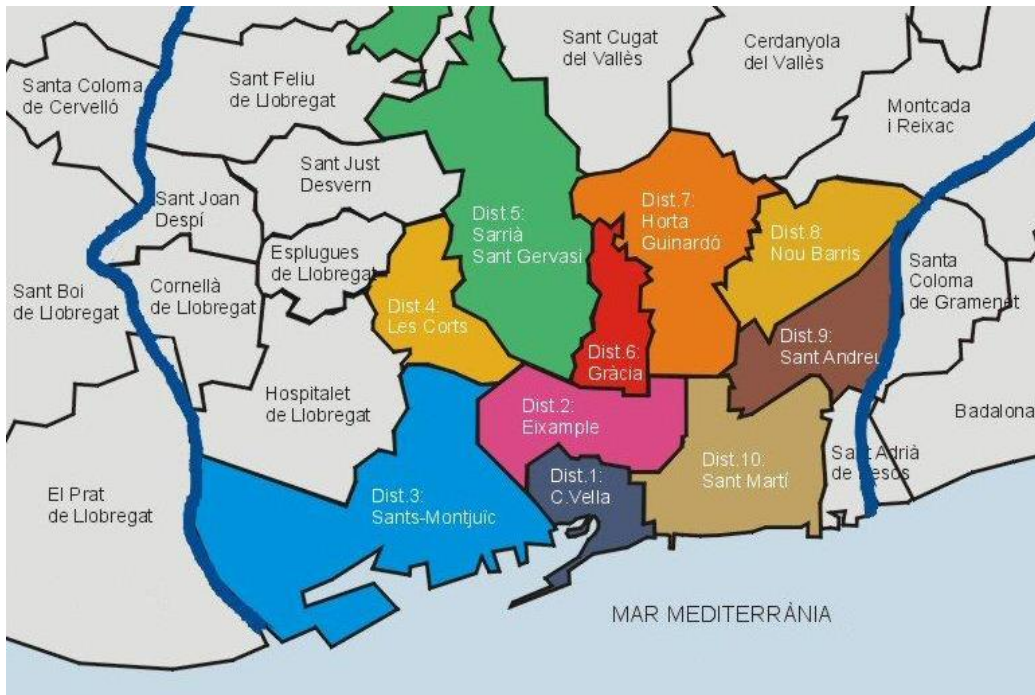


Figura 7.39: Mapa districtes de Barcelona.

Font: <https://mapabarcelona360.es/mapa-barrios-barcelona#.WxF9Y0iFM2x>

S'observa que els municipis d'origen es separen en dos grups: a l'esquerra els del sud de Barcelona i a la dreta els del nord (Figura 7.38) i la Figura 7.39 mostra la distribució real dels districtes de Barcelona. Ara s'identificaran les destinacions d'aquests municipis:

- Santa Coloma de Gramenet i Montcada i Reixac són municipis que estan molt relacionats amb els districtes 08019308 (Nou Barris) i 08019309 (Sant Andreu).

- Sant Adrià del Besòs està relacionat amb el districte 08019310 (Sant Martí).
- Tots els municipis del sud de Barcelona estan relacionats amb els districtes 08019302 (Eixample), 08019303 (Sants-Montjuïc), 08019304 (Les Corts), 08019305 (Sarrià, Sant Gervasi) i 08019311 (Pedralbes). Aquest últim districte és una codificació especial inclosa en la base de dades que ha proporcionat Kineo.

8. Conclusions

En aquest estudi es parteix de la premissa que la utilització dels mètodes de recollida de dades de transport i mobilitat mitjançant les noves tecnologies, té i ha de tenir encara més, un paper fonamental en el futur en la modelització del transport i l'estudi de patrons de mobilitat.

En aquest treball l'objectiu ha estat la validació de l'explotació de les dades procedents de la telefonia mòbil (elaborades per KINEO), via el contrast amb documents publicats per les autoritats de transport de l'Àrea Metropolitana de Barcelona (AMB), concretament amb els resultats de la darrera Enquesta de Mobilitat en Dia Feiner 2015 (EMEF 2015). Les dades explotades per KINEO han resultat tenir la consistència i qualitat necessària per a la modelització de la demanda de transport, horari i per franges, al Projecte Virtual Mobility LAB, malgrat alguns aspectes deficitaris. Per fer aquesta validació s'han definit uns indicadors a caracteritzar i contrastar amb l'EMEF, fluxos de viatges entre grans àmbits de la Regió Metropolitana de Barcelona (RMB) segmentats per mode de transport, motiu de viatge i lloc de residència del viatger.

Els indicadors de ratis de públic/privat i generació de viatges per persona indiquen una tendència a sota-estimar el nombre de viatges en l'explotació de les dades de telefonia mòbil. El problema de la sota-estimació no impedeix l'ús en modelització si l'estructura espai-temporal dels viatges és consistent amb la pràctica professional dels experts en transport de l'Àrea Metropolitana de Barcelona (AMB), com és en aquest cas. La metodologia emprada per validar la consistència espai-temporal ha estat l'anàlisi estadística avançada, per Components Principals (aplicat als patrons de viatges origen-destinació temporals) i Correspondències Simples (per validar la consistència de la distribució espacial dels patrons de viatges origen-destinació entre municipis, barris o districtes). Aquestes tècniques s'apliquen per primer cop en el sector del transport.

Respecte als resultats obtinguts de la validació referents a la RMB, amb les comparacions que s'han realitzat, la diferència entre el total de desplaçaments en un dia feiner són de 5 milions més en l'enquesta de l'EMEF, que segons les dades de KINEO Això és degut principalment a tres factors, que ja s'han explicat en detall, que són: les dades de Kineo no contemplen els desplaçaments entre els municipis de la RMB que no

pertanyen a la Primera Corona, per tant, no resulten directament contrastables, tampoc tenen en compte els residents a la resta de Catalunya (i Espanya) o als residents a l'estranger, pels quals l'EMEF aplica un factor de correcció que no es documenta.

En relació a la mobilitat modal, en les dades de KINEO s'observen moltes similituds, sobretot en el transport públic. Segons el motiu de viatge, no s'ha pogut comparar estrictament, ja que la definició dels motius no és la mateixa per KINEO, que per l'EMEF. Aquesta tendència s'observa tant en la validació dels patrons de la Regió Metropolitana com de la Primera Corona (PC).

Amb els resultats de la PC, hi ha una sensible diferència en el total dels desplaçaments diaris, els resultats de l'EMEF mostren uns 2 milions més que a les dades de Kineo, degut a dos dels factors explicats anteriorment (factor de correcció i viatges de no residents a la RMB).

La diferència més remarcable entre les dades de Kineo i EMEF es localitza al mode de transport, concretament en l'ús del transport públic en els viatges dins la PC excloent Barcelona, perquè és una diferència molt gran entre els dos resultats, els viatges de l'EMEF són quasi 8 vegades més que amb les dades de Kineo. Aquesta troballa té una explicació en el procés d'elaboració seguit: les dades elaborades per Kineo no disposaven del repartiment modal que fou estimat a la UPC (Montero et al. 2018) fusionant les dades de Kineo amb dades d'una enquesta antiga, EMIT (2006), de la qual es disposava del repartiment modal entre macrozones de la RMB (unes 100 macrozones). El nivell de detall de la informació modal entre macrozones és molt precari fora de la ciutat de Barcelona, a banda de molt obsolet (més de 10 anys).

L'anàlisi acurat de la distribució horària indica que les dades de Kineo comptabilitzen tots els desplaçaments i es poden veure 3 períodes marcats (matí, migdia i tarda-nit). Això és un punt a favor d'aquest tipus de dades ja que els experts en mobilitat pensen en les hores punta del dia en aquestes tres franges, i en canvi, en els resultats de l'EMEF només es remarquen els períodes punta de matí i tarda-nit i no es reflecteix la realitat. Aquest efecte és degut a que les dades d'EMEF procedeixen de l'expansió d'una enquesta obtinguda a partir d'una mostra domiciliària i en l'exploració dels detalls horaris, és ben coneguda la manca de fiabilitat dels resultats obtinguts.

L'aplicació de l'ACP en la planificació del transport és una novetat per validar la consistència temporal de les dades de viatges OD en el Projecte VML (Virtual Mobility Lab, (Barceló, Montero, and Ros-Roca 2018)). Els resultats de l'ACP en la projecció de les variables numèriques en els dos primers eixos principals són molt il·lustratius, ja que la variable latent associada al primer eix representa la quantitat de viatges i la del segon eix està relacionada amb la franja horària, indicant una bona consistència temporal.

La realització del clúster de l'ACP ha estat decebedora per dos motius: el primer, a causa del volum de les dades OD horàries, molts algoritmes no deixaven fer els càlculs perquè demanen una memòria de treball (memòria RAM) molt elevada, no disponible; el segon, després de la realització de la classificació a partir de l'algoritme del K-means, no

s'ha pogut trobar cap patró que pugui caracteritzar clarament cadascun dels clústers. Per tant, sense dades sociodemogràfiques addicionals, no disponibles en aquest treball, no s'ha pogut extraure una classificació entenedora dels patrons OD horaris inclosos a les dades de Kineo.

L'ACS, per la validació detallada de la distribució espacial dels viatges, ha permès en alguns casos una caracterització dels perfils OD espacials sorprenents. Aquest mètode ha revelat patrons de mobilitat amb molta informació sobre les dades, per tant aquesta anàlisi és una bona opció per emprar en validacions futures de consistència espacial en patrons de mobilitat. Algunes troballes sorprenents han estat:

- En els perfils de mobilitat diària, global i per modes, entre els barris de Barcelona, on s'observa clarament la geografia dels barris de Barcelona,
- En la distribució de l'intercanvi de viatges entre municipis de la PC segons la seva proximitat, i ubicant la ciutat de Barcelona com el centre de gravetat,.
- En la distribució de la mobilitat entre municipis de la PC i els districtes municipals de Barcelona-ciutat.

Com futures investigacions, per assolir una metodologia sistemàtica per la validació de consistència de patrons de mobilitat s'entreveuen diverses opcions:

- La primera seria la utilització de la interfície Spark mitjançant R-Studio, per treballar al núvol (internet) amb grans conjunts de dades i a partir d'aquí realitzar un altre cop l'ACP, una classificació no dirigida a partir de l'ACP i l'ACS.
- La segona seria incorporar dades socioeconòmiques de les zones de transport i aplicar una Anàlisi de Correspondències Múltiples (ACM).
- Finalment, aprofundir en el calibratge dels models de distribució, per motius i per modes, importants pel model UTPS a desenvolupar al Projecte VML.

En relació amb la consecució dels objectius identificats inicialment, es pot afirmar que s'ha assolit el desenvolupament d'una estratègia d'anàlisi i validació espai-temporal de dades relatives a patrons de mobilitat amb la finalitat d'emprar les dades en entorns de planificació estratègica dels sistemes de transport segons el marc conceptual UTPS.

La validació efectuada contribueix, en una fase prèvia, a descartar els dubtes sobre la utilització de les dades sobre patrons de mobilitat de Kineo en el Projecte VML.

L'estratègia d'anàlisi desenvolupada en aquest treball es podrà aplicar a futures trameses com les relacionades amb les dades de mobilitat de INRIX que estan a punt de posar-se en marxa durant el 2018 a l'inLab.

Els resultats d'aquesta estratègia d'anàlisi i les conclusions relatives a la validesa de les dades de Kineo s'han presentat al Workshop anual de Kineo a Madrid (23 Maig 2018, Madrid).

Aquest ha estat el primer projecte de gran calibre que he hagut de desenvolupar des de zero relacionat amb l'àmbit empresarial i m'ha permès aprendre tot el procés que s'ha de

seguir per realitzar una proposta inicial i planificació i el seu posterior desenvolupament d'un treball. També m'ha permès millorar professionalment perquè he tingut l'oportunitat de interactuar directament amb diferents persones involucrades a l'àmbit de la planificació del transport.

Així mateix l'haver hagut de presentar els resultats en el Workshop de KINEO m'ha ajudat a aprendre l'altra banda del desenvolupament d'un projecte: haver d'exposar els resultats obtinguts a un públic, tant d'expert en l'àrea, com als que no. Finalment, comentar que si hagués tingut més temps, m'hauria agradat emprar més eines analítiques que en el futur podessin ser útils en el projecte VML.

El present treball m'ha donat l'oportunitat d'aprofundir en assignatures de Data Science que havia estudiat al Grau d'Estadística com l'Anàlisi Multivariant de Dades i Minería de Dades. Tanmateix, l'assignatura Inferència Estadística Avançada del MESIO m'ha permès entendre amb facilitat el marc de modelització i estimació de models UTPS, sobretot en l'aspecte dels models de distribució i repartiment modal de viatges i l'assignatura Models Discrets en Xarxes que m'ha permès entendre la problemàtica de la selecció de rutes que apareix en els problemes d'assignació de transport.

Bibliografia

- Autoritat del Transport Metropolità. 2015. “Enquesta de Mobilitat En Dia Feiner 2015 AMB.” 2015(Emef).
- Backstrom, Lars, Eric Sun, and Cameron Marlow. 2010. “Find Me If You Can : Improving Geographical Prediction with Social and Spatial Proximity.” *Proceeding WWW '10 Proceedings of the 19th international conference on World wide web*: 61–70.
- Barceló, J., L. Montero, and X. Ros-Roca. 2018. “Virtual Mobility LAB.” <http://hdl.handle.net/2117/113344>.
- Chen, Cynthia et al. 2016. “The Promises of Big Data and Small Data for Travel Behavior (Aka Human Mobility) Analysis.” *Transportation Research Part C: Emerging Technologies* 68: 285–99. <http://dx.doi.org/10.1016/j.trc.2016.04.005>.
- Cheng, Zhiyuan, James Caverlee, and Kyumin Lee. 2010. “You Are Where You Tweet : A Content-Based Approach to Geo-Locating Twitter Users.” *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*: 759–68.
- Cheng, Zhiyuan, James Caverlee, Kyumin Lee, and Daniel Z. Sui. 2011. “Exploring Millions of Footprints in Location Sharing Services.” *Icwsn 2010(Cholera)*: 81–88.
- Golder, S. A., D. Wilkinson, and B. A. Huberman. 2007. “Rhythms of Social Interaction: Messaging Within.” *Proceedings of the Third Communities and Technologies Conference*: 41–66. https://books.google.com.br/books?hl=en&lr=&id=HFfa7AeNIYtUC&oi=fnd&pg=PA40&ots=wUFn-m-_1L&sig=Q3g6c_uwL_0RyydKWLHHilxLZzk.
- Husson, François, Sébastien Lê, and Jérôme Pagès. 2011. *Journal of Applied Statistics Exploratory Multivariate Analysis by Example Using R*. <http://www.tandfonline.com/doi/abs/10.1080/02664763.2012.657409>.
- Lindqvist, Janne et al. 2011. “I’m the Mayor of My House: Examining Why People Use Foursquare - a Social-Driven Location Sharing Application.” *CHI '11 Proceedings of the 2011 annual conference on Human factors in computing systems* 54(6): 2409–18. <http://www.cs.cmu.edu/~jwwiese/publications/chi2011-foursquare.pdf>.
- Long, Ying, and Jean-claude Thill. 2015. “Combining Smart Card Data and Household Travel Survey to Analyze Jobs-Housing Relationships in Beijing.Pdf.”

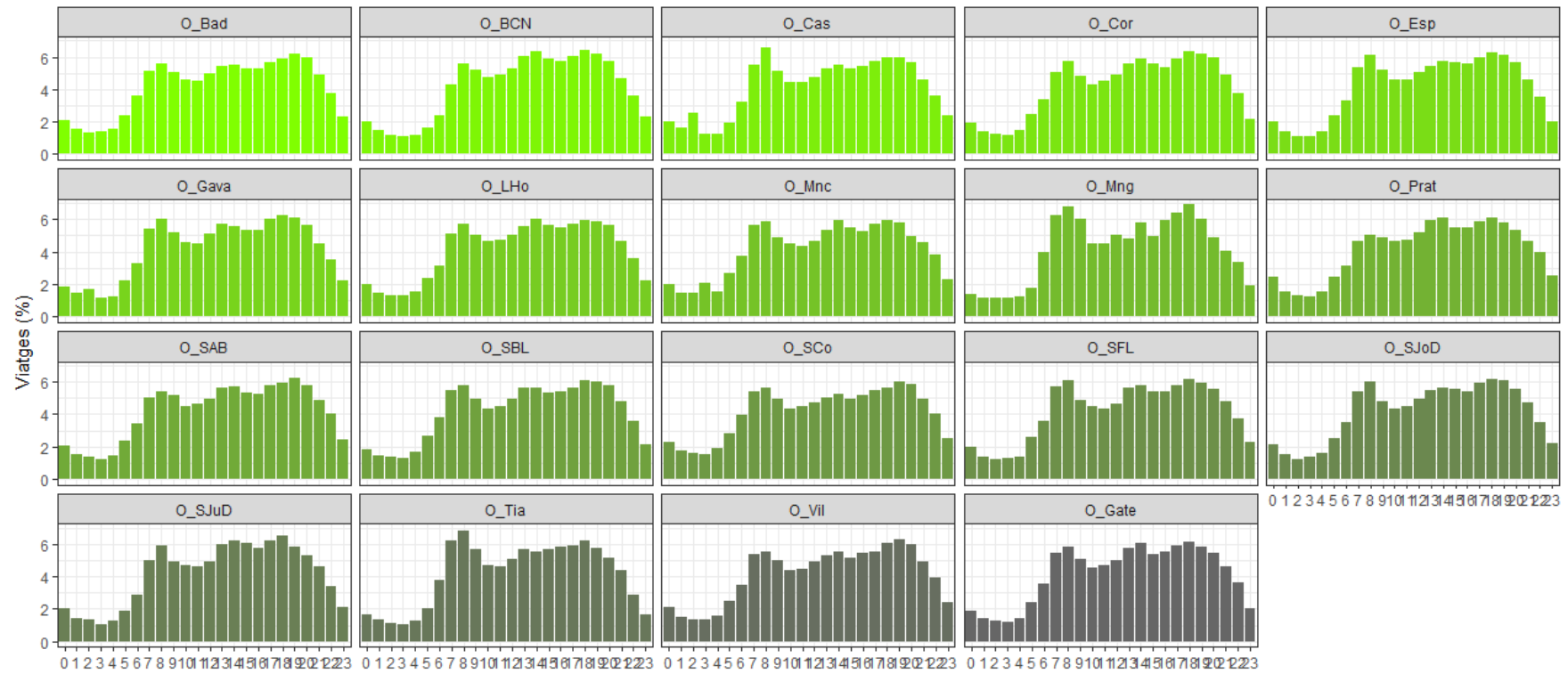
- Montero, L, X Ros-roca, R Herranz, and J Barceló. 2018. "Fusing Mobile Phone Data with Other Data Sources to Generate Input OD Matrices for Transport Models." (September): 17–19.
- Ortuzar, Juan de Dios, and Luis G. Willumsen. 2011. *Modelling Transport Modelling Transport*.
- Plummer, Andrew V. 1962. "Creating the First Plan (1955-1962) By."
- Yardi, Sarita, and Danah Boyd. 2010. "Tweeting from the Town Square : Measuring Geographic Local Networks." *Methods* 77(3): 194–201.
<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1490/1853>.
- Zhang, Faming et al. 2016. "Analyzing Urban Human Mobility Patterns through a Thematic Model at a Finer Scale." *ISPRS International Journal of Geo-Information* 5(6): 78. <http://www.mdpi.com/2220-9964/5/6/78>.

Apèndix

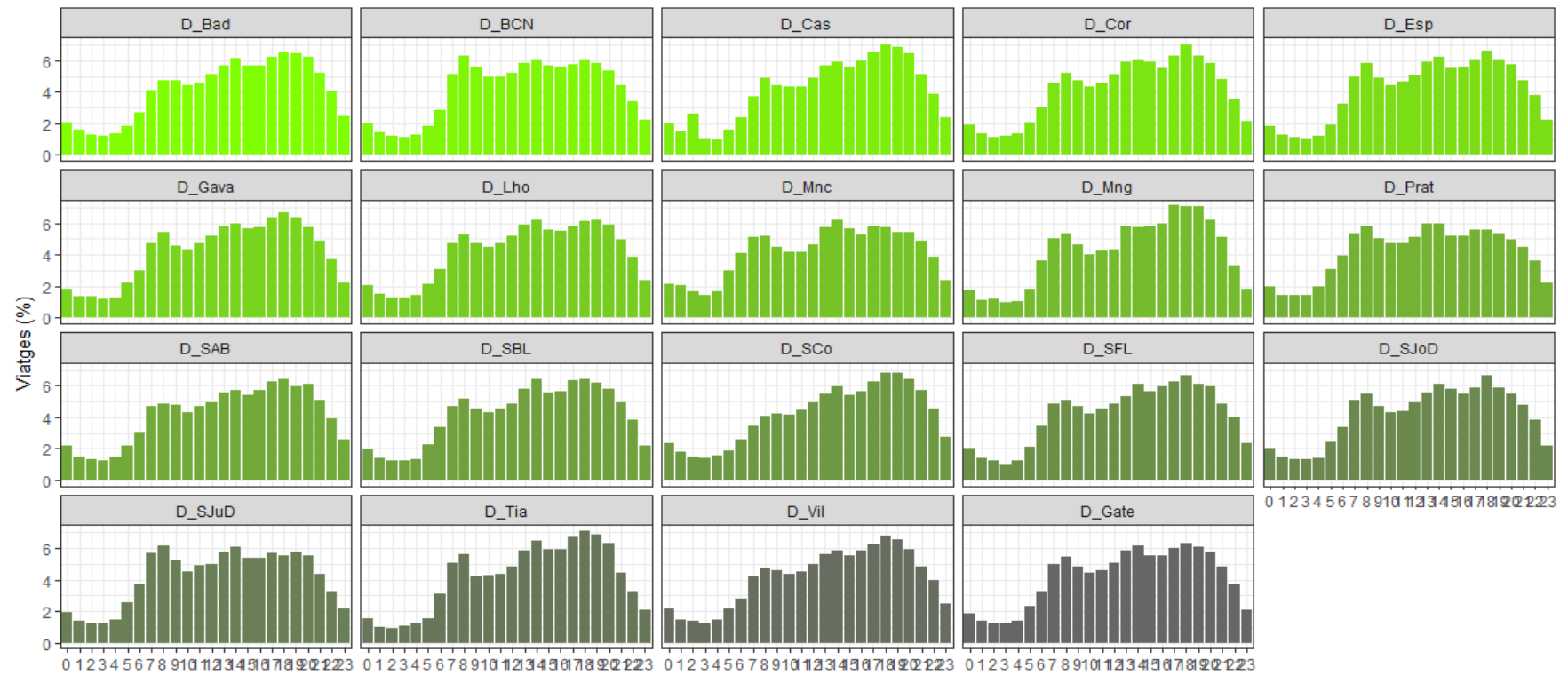
A. Matriu de viatges origen-destinació

Municipis	Badalona	Barcelona	Castelldefels	Cornellà de Llobregat	Esplugues de Llobregat	Gavà	L'Hospitalet de Llobregat	Montcada i Reixac	Montgat	El Prat de Llobregat	Sant Adrià de Besòs	Sant Boi de Llobregat	Sant Feliu de Llobregat	Sant Joan Despí	Sant Just Desvern	Santa Coloma de Gramenet	Tiana	Viladecans	Total
Badalona	286.918	111.719	343	1.119	1.037	415	3.609	5.972	3.866	1.484	32.650	701	186	323	463	43.963	5.106	405	500.279
Barcelona	111.803	4.288.894	14.607	38.045	37.173	13.643	204.606	27.269	4.044	54.269	52.600	24.382	9.522	12.636	11.738	93.694	3.620	14.203	5.016.747
Castelldefels	396	14.969	60.987	1.529	722	22.868	3.201	90	26	3.428	93	3.392	321	410	287	154	20	8.662	121.553
Cornellà de Llobregat	1.100	38.741	1.589	74.231	18.513	1.733	34.811	340	67	7.105	391	12.522	3.711	15.349	5.664	488	51	2.616	219.020
Esplugues de Llobregat	945	37.202	721	19.129	25.148	913	32.593	322	64	2.255	267	2.899	1.816	5.358	9.727	430	64	1.220	141.073
Gavà	397	13.630	23.659	1.752	834	32.138	3.872	116	27	5.429	64	4.900	434	536	314	160	26	22.810	111.097
L'Hospitalet de Llobregat	3.597	203.725	3.218	35.534	32.397	3.999	287.535	1.041	158	14.573	1.113	9.418	2.896	4.126	3.028	1.670	118	5.230	613.376
Montcada i Reixac	6.105	26.110	103	325	379	115	898	21.643	116	329	830	248	101	106	86	6.187	143	117	63.941
Montgat	3.801	3.523	18	40	77	30	136	144	1.037	25	354	22	10	25	14	499	977	10	10.743
El Prat de Llobregat	1.470	53.095	3.490	6.828	1.972	5.514	14.307	316	43	89.375	419	10.127	1.249	1.837	733	773	58	7.640	199.245
Sant Adrià de Besòs	32.994	53.699	89	339	220	84	1.095	653	315	497	13.801	226	79	71	84	4.040	351	124	108.759
Sant Boi de Llobregat	746	24.650	3.489	12.075	2.747	5.015	9.183	207	19	9.988	198	95.035	5.998	8.757	3.010	351	45	12.534	194.047
Sant Feliu de Llobregat	210	10.676	298	3.882	1.923	495	3.036	86	6	1.482	93	5.922	19.867	5.843	7.723	107	15	643	62.296
Sant Joan Despí	366	12.795	470	15.258	5.417	553	3.861	125	25	1.901	102	8.909	6.200	8.222	7.008	173	9	832	72.226
Sant Just Desvern	414	11.074	260	5.822	9.523	324	3.103	85	13	826	116	3.104	7.821	7.341	5.574	169	16	388	55.972
Santa Coloma de Gramenet	43.977	93.609	139	490	631	174	1.803	5.903	557	783	3.959	332	90	160	203	98.626	646	166	252.248
Tiana	5.160	3.376	23	39	84	14	124	151	1.098	62	377	38	5	15	22	642	919	22	12.171
Viladecans	438	14.521	8.657	2.734	1.169	22.727	5.074	165	20	7.354	109	12.194	614	884	448	196	35	66.842	144.183
Total	500.838	5.016.008	122.160	219.169	139.965	110.744	612.850	64.629	11.500	201.165	107.537	194.369	60.918	71.998	56.125	252.321	12.220	144.461	7.898.975

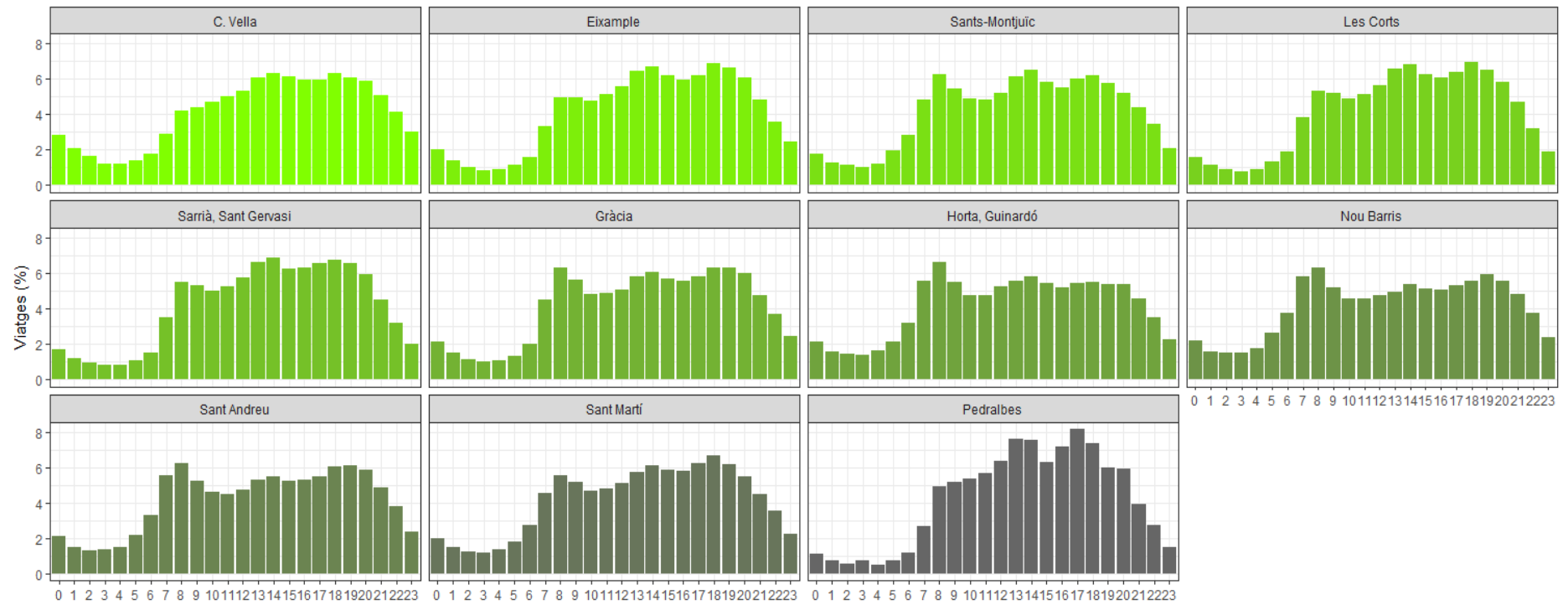
B. Franja horària per municipi origen



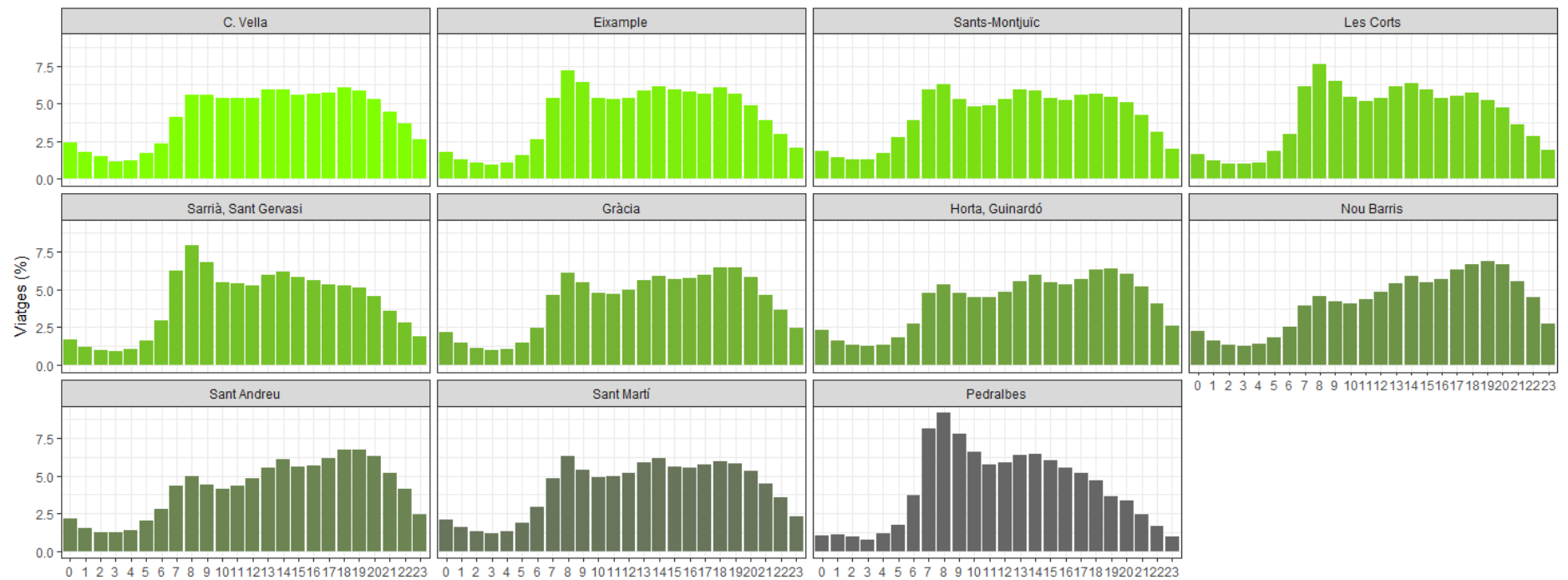
C. Franja horària per municipi destinació



D. Franja horària per districte origen de Barcelona



E. Franja horària per districte destinació de Barcelona



F. Codi R

Només hi ha un fragment del codi, que fa referencia al estudi general de la RMB i la PC, aquest codi mitjançant unes modificacions s'aplica a les matrius OD segons el motiu del viatge i segons el lloc de residència.

```
# Paquets que són necessaris
requiredPackages <- c("knitr", "AER", "VGAM", "mlogit", "lmtest",
"effects", "FactoMineR", "car",
"factoextra", "RColorBrewer", "ggplot2", "corrplot", "pander", "plyr", "resh
ape2", "cluster", "mvoutliers", "scales")
missingPackages <- requiredPackages[!(requiredPackages %in%
installed.packages()[, "Package"])]

if(length(missingPackages)) install.packages(missingPackages)
lapply(requiredPackages, require, character.only = TRUE)

#Llegim dades kineo segona entrega
setwd("F:/KINEOOD_V2")
df<-
read.table("ODEMIT_KINEO_v2.csv", header=T, sep=";", check.names=T, quote
= "")

#####
#####DADES#####
#####
setwd("F:/KINEOOD_V3/res/noves")
baseAll<-read.csv("matrix_all_all.csv", header=T, check.names=T)
baseAll <- baseAll[,-c(1,48:49)]

#Calculem les següents columnes de l'entrega 3 a partir dels modal
split de l'entrega 2
baseAll$dm_kineo_car<- 0.01*df$ems_car*baseAll$tot_kineo
baseAll$dm_kineo_bus<- 0.01*df$ems_bus*baseAll$tot_kineo
baseAll$dm_kineo_met<- 0.01*df$ems_met*baseAll$tot_kineo
baseAll$dm_kineo_other<- 0.01*df$ems_other*baseAll$tot_kineo

#####
#####PRE-PROCESSAMENT#####
#####
#Donem un cop d'ull a les dades
str(baseAll)
summary(baseAll)

#####
#Les observacions NA de les següents variables les anomenarem portes
ll<-which(is.na(baseAll$nombarori));length(ll)
llev<-c(levels(baseAll$nombarori), "Barri_Gate")
dllev<-c(levels(baseAll$nombardes), "Barri_Gate")
table(baseAll$nombarori, useNA="always")
```



```

baseAll$nombarori<-as.numeric(baseAll$nombarori)
baseAll[11,"nombarori"]<-73
11<-which(is.na(baseAll$nombardes));length(11)
baseAll$nombardes<-as.numeric(baseAll$nombardes)
baseAll[11,"nombardes"]<-73
baseAll$nombarori<-factor(baseAll$nombarori,labels=1lev)
baseAll$nombardes<-factor(baseAll$nombardes,labels=dllev)

11<-which(is.na(baseAll$comori));length(11)
1lev<-c(levels(baseAll$comori),"Com_Gate")
dllev<-c(levels(baseAll$comdes),"Com_Gate")
table(baseAll$comori,useNA="always")
baseAll$comori<-as.numeric(baseAll$comori)
baseAll[11,"comori"]<-5
11<-which(is.na(baseAll$comdes));length(11)
baseAll$comdes<-as.numeric(baseAll$comdes)
baseAll[11,"comdes"]<-5
baseAll$comori<-factor(baseAll$comori,labels=1lev)
baseAll$comdes<-factor(baseAll$comdes,labels=dllev)

11<-which(is.na(baseAll$munos));length(11)
1lev<-c(levels(baseAll$munos),"O_Gate")
dllev<-c(levels(baseAll$munds),"D_Gate")
table(baseAll$munos,useNA="always")
baseAll$munos<-as.numeric(baseAll$munos)
baseAll[11,"munos"]<-19
11<-which(is.na(baseAll$munds));length(11)
baseAll$munds<-as.numeric(baseAll$munds)
baseAll[11,"munds"]<-19
baseAll$munos<-factor(baseAll$munos,labels=1lev)
baseAll$munds<-factor(baseAll$munds,labels=dllev)

#Comprovem que ja no hi ha NA's
summary(baseAll)

#####
#####MATRIU OD-PC#####
#####

Pobmatriu <- matrix(rep(0,18*18), ncol=18,nrow=18)
Pobmatriu <- data.frame(Pobmatriu)
noms <- levels(baseAll$munori)
noms <- noms[-6]
colnames(Pobmatriu) <- noms
row.names(Pobmatriu)<- noms
for(i in 1:18){
  for(j in 1:18){
    Pobmatriu[i,j] <- sum(baseAll$tot_kineo[baseAll$munori==noms[i] &
baseAll$mundes==noms[j]])
  }
}

#####
#####FRANGES HORARIES#####
#####
#####
#Total
#####
hores <- NULL
hores2<-NULL
j=1

```

```

for(i in 7:30){
  hores[j] <- round((sum(baseAll[,i])/sum(baseAll$tot_kineo))*100,2)
  hores2[j] <- sum(baseAll[,i])
  j <- j+1
}

hores <- data.frame(hores=c(0:23), viatges=hores,
group=c(rep(0,7),1,1,1,0,0,0,1,1,0,0,1,1,1,0,0,0,0))

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-1
color <- colours[it]
colfunc <- colorRampPalette(c("grey40",color))
colours_pal <- colfunc(nlevels(as.factor(hores$group)))
names(colours_pal) <- levels(as.factor(hores$group))

g1<- ggplot(hores,aes(x= as.factor(hores), y= viatges,
fill=as.factor(group)))
gg1<-g1+geom_bar(stat='identity', color='black')+theme_classic()+
  ggtitle('Franja horaria') + xlab("Hores")+ylab("Viatges (%)")+
  theme(axis.line.x = element_blank(),axis.line.y = element_blank(),
        plot.title = element_text(hjust = 0.5,size=15),
        axis.title.x = element_text(size = 13), axis.title.y =
element_text(size = 13))
gg1+ scale_fill_manual(values=colours_pal, guide=F)

#####
#Municipi origen
#####
nommun <- levels(baseAll$munos)
nn <- (length(nommun)+1)*24
Hormun <- matrix(rep(0,nn), ncol=20,nrow=24)
Viatmun <- data.frame(Hormun)
names(Viatmun)<- c('Hora', nommun)
l1 <- c('Hora', nommun)
colnames(Viatmun)<- l1
Viatmun[,1]<- as.factor(c(0:23))
l <- 2
for(i in 1:(length(nommun))){
  k <- 1
  for(j in 7:30){
    sel <- which(baseAll$munos==nommun[i])
    Viatmun[k,1] <-
as.numeric(round((sum(baseAll[sel,j])/sum(baseAll$tot_kineo[sel]))*100
,2))
    k <- k+1
  }
  l <- l+1
}

Viatmun2 <- melt(Viatmun, id.var="Hora")

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(Viatmun2$variable))
names(colours_pal) <- levels(Viatmun2$variable)

```

```

g2<-ggplot(Viatmun2, aes(x=as.factor(Hora),y=value))+
  geom_bar(stat='identity',aes(fill=as.factor(variable)))
g2 + facet_wrap(~variable)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+
  ggtitle("Franja horària per municipi origen (%)")+theme(plot.title =
element_text(hjust = 0.5))

#####
#Municipi destinació
#####
nommun <- levels(baseAll$munds)
nn <- (length(nommun)+1)*24
Hormun <- matrix(rep(0,nn), ncol=20,nrow=24)
Viatmun <- data.frame(Hormun)
names(Viatmun)<- c('Hora', nommun)
ll <- c('Hora', nommun)
colnames(Viatmun)<- ll
Viatmun[,1]<- as.factor(c(0:23))
l <- 2
for(i in 1:(length(nommun))){
  k <- 1
  for(j in 7:30){
    sel <- which(baseAll$munds==nommun[i])
    Viatmun[k,l] <-
as.numeric(round((sum(baseAll[sel,j])/sum(baseAll$tot_kineo[sel]))*100
,2))
    k <- k+1
  }
  l <- l+1
}

Viatmun2 <- melt(Viatmun, id.var="Hora")

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(Viatmun2$variable))
names(colours_pal) <- levels(Viatmun2$variable)

g3<-ggplot(Viatmun2, aes(x=as.factor(Hora),y=value))+
  geom_bar(stat='identity',aes(fill=as.factor(variable)))
g3 + facet_wrap(~variable)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+
  ggtitle("Franja horària per municipi destinació
(%")+theme(plot.title = element_text(hjust = 0.5))

#####
#Districte a Barcelona Origen
#####
sel<-
which((baseAll$nombarori=="Barri_Gate")|(baseAll$nombardes=="Barri_Gat
e'))
dfddistricte<-baseAll[-sel,]
nomdis <- levels(factor(dfddistricte$mundisori))
nn <- (length(nomdis)+1)*24
Hordis <- matrix(rep(0,nn), ncol=12,nrow=24)

```

```

Viatdis <- data.frame(Hordis)
names(Viatdis)<- c('Hora', nomdis)
ll <- c('Hora', nomdis)
colnames(Viatdis)<- ll
Viatdis[,1]<- as.factor(c(0:23))
l <- 2
for(i in 1:(length(nomdis))){
  k <- 1
  for(j in 7:30){
    sel <- which(baseAll$mundisori==nomdis[i])
    Viatdis[k,1] <-
as.numeric(round((sum(baseAll[sel,j])/sum(baseAll$tot_kineo[sel]))*100
,2))
    k <- k+1
  }
  l <- l+1
}

Viatdis2 <- melt(Viatdis, id.var="Hora")

levels(Viatdis2$variable) <- c('C. Vella', 'Eixample','Sants-
Montjuïc',
                                'Les Corts','Sarrià, Sant Gervasi',
'Gràcia',
                                'Horta, Guinardó','Nou Barris','Sant
Andreu',
                                'Sant Martí','Pedralbes')

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(Viatdis2$variable))
names(colours_pal) <- levels(Viatdis2$variable)

g4<-ggplot(Viatdis2, aes(x=as.factor(Hora),y=value))+
  geom_bar(stat='identity',aes(fill=as.factor(variable)))
g4 + facet_wrap(~variable)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+
  ggtitle("Franja horària per districte de Barcelona origen
(%)" )+theme(plot.title = element_text(hjust = 0.5))

#####
#Districte a Barcelona Destinació
#####
nn <- (length(nomdis)+1)*24
Hordis <- matrix(rep(0,nn), ncol=12,nrow=24)
Viatdis <- data.frame(Hordis)
names(Viatdis)<- c('Hora', nomdis)
ll <- c('Hora', nomdis)
colnames(Viatdis)<- ll
Viatdis[,1]<- as.factor(c(0:23))
l <- 2
for(i in 1:(length(nomdis))){
  k <- 1
  for(j in 7:30){
    sel <- which(baseAll$mundisdes==nomdis[i])

```

```

    Viatdis[k,l] <-
as.numeric(round((sum(baseAll[sel,j])/sum(baseAll$tot_kineo[sel]))*100
,2))
    k <- k+1
  }
  l <- l+1
}

Viatdis2 <- melt(Viatdis, id.var="Hora")

levels(Viatdis2$variable) <- c('C. Vella', 'Eixample','Sants-
Montjuïc',
                                'Les Corts','Sarrià, Sant Gervasi',
'Gràcia',
                                'Horta, Guinardó','Nou Barris','Sant
Andreu',
                                'Sant Martí','Pedralbes')

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(Viatdis2$variable))
names(colours_pal) <- levels(Viatdis2$variable)

g5<-ggplot(Viatdis2, aes(x=as.factor(Hora),y=value))+
  geom_bar(stat='identity',aes(fill=as.factor(variable)))
g5 + facet_wrap(~variable)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+
  ggtitle("Franja horària per districte de Barcelona destinació
(%)")+theme(plot.title = element_text(hjust = 0.5))

#####
#####TOTAL VIATGES SEGONS MODE DE TRANSPORT#####
#####
Transport <-
matrix(c("Cotxe","Bus","Metro","Altres","Total",sum(baseAll$dm_kineo_c
ar),sum(baseAll$dm_kineo_bus),sum(baseAll$dm_kineo_met),sum(baseAll$dm
_kineo_other),sum(baseAll[,43:46])),ncol=2)

a<-NULL
for(i in 1:4){
  a[i]<-
round((as.numeric(Transport[i,2])/as.numeric(Transport[5,2]))*100,2)
}

Transport2 <- data.frame(rep(1,4),Transport[-5,], a)
colnames(Transport)<- c("Tipus Transport","viatges Realitzats")
colnames(Transport2)<-
c("X","TipusTransport","viatgesRealitzats","Percentatgeviatges")

Transport2$viatgesRealitzats<-
round(as.numeric(as.character(Transport2$viatgesRealitzats)),2)
pandoc.table(Transport2[,-1])

Transport2 <- ddply(Transport2, "X", transform, Etiquetes =
paste(round( Percentatgeviatges, digits=2), '%'))
Transport2 <- ddply(Transport2, "X", transform, PosEtiq =
c(22,40,4.5,70))

```

```

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(Transport2$TipusTransport))
names(colours_pal) <- levels(Transport2$TipusTransport)

g6<-ggplot(Transport2,
aes(x="",y=Percentatgeviatges,fill=as.factor(TipusTransport)))
g6+geom_bar(stat='identity', width=1, color='black', size=0.7,position
= position_stack()) +
  scale_fill_manual(values=colours_pal,name = "")+
  geom_text(aes(y=PosEtiq,label=Etiquetes), size=5,col="white")+
theme_classic()+
  theme(axis.line.x = element_blank(),axis.line.y = element_blank(),
        axis.text.x=element_text(color='black',size=12),legend.text =
element_text(size=12),
        legend.position = "bottom")+
  coord_polar("y") +xlab("")+ylab("") +
  scale_y_continuous(breaks=NULL,labels=FALSE)

g6<-ggplot(Transport2, aes(x="",y=Percentatgeviatges))+
  geom_bar(stat='identity',aes(fill=TipusTransport))
g6 +
  facet_grid(~TipusTransport)+scale_fill_manual(values=colours_pal)+
  theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+
  ggtitle("Mode de transport (%)")+theme(plot.title =
element_text(hjust = 0.5))

#####
#####ÀREA DE MOBILITAT SEGONS KINEO#####
#####
#Selecció de les pitjors observacions de les variables ori i des
sel<-which((baseAll$ori>710)|(baseAll$des>710))
if (length(sel)==0) {
  dfi<-baseAll
} else {
  dfi<-baseAll[-sel,]
}
#Seleccionar els orogens que provenen de l'exterior
from.rm2pc<-which(((dfi$ori>699)&(dfi$ori<711))&(dfi$des<700))
#Seleccionar els destins exteriors
from.pc2rm<-which(((dfi$ori<700)&(dfi$des>699)&(dfi$des<711)))
#Seleccionar viatges en un dia laborable dins la primera corona a RMB
intpc<-which((dfi$ori<700)&(dfi$des<700))
#Seleccionar viatges en un dia laborable dins RMB, fora de la primera
corona
intrm<-which(((dfi$ori>699))&(dfi$des>699))

zones <-
data.frame(Zona=c('IntraPC','RMBnoPC','Commuters','Total'),viatges=rep
(1,4),Percenviatges=rep(1,4))
#Total viatges en un dia laborable a RMB
zones[4,2]<-sum(dfi$tot_kineo)
zones[4,3]<- 100
#Total viatges en un dia laborable dins la primera corona a RMB
zones[1,2]<-sum(dfi$tot_kineo[intpc])
zones[1,3]<-(sum(dfi$tot_kineo[intpc])/sum(dfi$tot_kineo))*100

```

```
#Total viatges en un dia laborable de/a RMB
zones[3,2]<-sum(dfi$tot_kineo[c(from.rm2pc,from.pc2rm)])
zones[3,3]<-
(sum(dfi$tot_kineo[c(from.rm2pc,from.pc2rm)]/sum(dfi$tot_kineo))*100
#Total viatges en un dia laborable dins RMB, fora de la primera corona
zones[2,2]<-sum(dfi$tot_kineo[intrm])
zones[2,3]<-(sum(dfi$tot_kineo[intrm])/sum(dfi$tot_kineo))*100

#zones <- ddpily(zones, 'Percenviatges', transform, Percenviatges =
paste(round( Percenviatges, digits=1), '%'))

pandoc.table(zones)

dfi$con<-1
dfi$con[intpc]<-2
dfi$con[c(from.rm2pc,from.pc2rm)]<-3

dfi$con<-factor(dfi$con,labels=c("RMBnoPC","IntraPC","Commuters"))
xtabs(dfi$tot_kineo~dfi$con)/sum(dfi$tot_kineo)
table(dfi$con,useNA="ifany")

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-2
color <- colours[it]
colfunc <- colorRampPalette(c(color,'grey40'))
colours_pal <- colfunc(nlevels(zones$Zona))
names(colours_pal) <- levels(zones$Zona)

g7<-ggplot(zones[-4,], aes(x="",y=Percenviatges))+
  geom_bar(stat='identity',aes(fill=Zona))
g7 + facet_grid(~Zona)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+
  ylab('Viatges (%)')+ guides(fill=FALSE)+ ggtitle("Àrea de
mobilitat")+
  theme(plot.title = element_text(hjust = 0.5,size=15),axis.title.y =
element_text(size = 15))

#####
#####PRIMERA CORONA SEGONS KINEO#####
#####
#Seleccionar aquells que provenen de l'exterior
sel<-which((baseAll$ori>699)|(baseAll$des>699))
#Eleminar les observacions seleccionades
dffc<-baseAll[-sel,]

#Seleccionar origen Barcelona
from.bcn2pc<-which(((dffc$ori>299)&(dffc$ori<700))&(dffc$des<300))
#Seleccionar origen Primera Corona
from.pc2bcn<-which(((dffc$ori<300)&((dffc$des>299)&(dffc$des<700)))
#Seleccionar viatges dins de Barcelona
intbcn<-which(((dffc$ori>299)&(dffc$des>299))
#Seleccionar viatges dins la Primera Corona
intpc<-which(((dffc$ori<300))&(dffc$des<300))

zonesPC <-
data.frame(Zona=c('IntraBCN','IntraPC','Commuters','Total'),viatges=rep
p(1,4),Percenviatges=rep(1,4))

#Total viatges en un dia laborable en la Primera Corona en la RMB
zonesPC[4,2]<-sum(dffc$tot_kineo)
```

```

zonesPC[4,3]<-100
#Total viatges en un dia laborable dins la Primera Corona
zonesPC[2,2]<-sum(dfffc$tot_kineo[intpc])
zonesPC[2,3]<-(sum(dfffc$tot_kineo[intpc])/sum(dfffc$tot_kineo))*100
#Total viatges en un dia laborable de BCN a PC o visceversa
zonesPC[3,2]<-sum(dfffc$tot_kineo[c(from.bcn2pc,from.pc2bcn)])
zonesPC[3,3]<-
  (sum(dfffc$tot_kineo[c(from.bcn2pc,from.pc2bcn)])/sum(dfffc$tot_kineo))*
  100
#Total viatges en un dia laborable dins de Barcelona
zonesPC[1,2]<-sum(dfffc$tot_kineo[intbcn])
zonesPC[1,3]<-(sum(dfffc$tot_kineo[intbcn])/sum(dfffc$tot_kineo))*100

pandoc.table(zonesPC)

dfffc$con<-1
dfffc$con[intbcn]<-2
dfffc$con[c(from.bcn2pc,from.pc2bcn)]<-3

dfffc$con<-factor(dfffc$con,labels=c("IntraFC","IntraBCN","Commuters"))
xtabs(dfffc$tot_kineo~dfffc$con)
cc<-xtabs(dfffc$tot_kineo~dfffc$con)/sum(dfffc$tot_kineo);cc
table(dfffc$con,useNA="ifany")

zonesPC1 <- zonesPC[-4,]
zonesPC1 <- ddply(zonesPC1, "Percenviatges", transform, X = 1)
zonesPC1 <- ddply(zonesPC1, 'X', transform, Etiquetes = paste(round(
Percenviatges, digits=1), '%'))
zonesPC1 <- ddply(zonesPC1, "X", transform, PosEtiq = c(3,94,45))
zonesPC1 <- zonesPC1[,-4]

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-2
color <- colours[it]
colfunc <- colorRampPalette(c(color,'grey40'))
colours_pal <- colfunc(nlevels(zonesPC1$Zona))
names(colours_pal) <- levels(zonesPC1$Zona)

g8<-ggplot(zonesPC1[-4,], aes(x="",y=Percenviatges))+
  geom_bar(stat='identity',aes(fill=Zona))
g8 + facet_wrap(~Zona)+scale_fill_manual(values=colours_pal)+
  theme_bw()+xlab("")+
  ylab('Viatges (%)')+guides(fill=FALSE)+ ggtitle("Primera Corona")+
  theme(plot.title = element_text(hjust = 0.5,size=15),axis.title.y =
  element_text(size = 15))

#####
#####PRIMERA CORONA SEGONS MODE DE TRANSPORT#####
#####
#S'utilitza la mateixa base de dades de l'apartat anterior
#dfffc

zonesPCMS <-
data.frame(Zona=c('IntraBCN','IntraPC','Commuters','Total'),Viatges=rep(1,4),Cotxe=rep(1,4),Bus=rep(1,4),Metro=rep(1,4),Altres=rep(1,4))
zonesPCMS2 <-
data.frame(Zona=c('IntraBCN','IntraPC','Commuters'),Cotxe=rep(1,3),Bus=rep(1,3),Metro=rep(1,3),Altres=rep(1,3))

#Total viatges PC

```



```

zonesPCMS[4,2]<- sum(dffc[,43:46])
zonesPCMS[4,3]<- sum(dffc$dm_kineo_car)
zonesPCMS[4,4]<- sum(dffc$dm_kineo_bus)
zonesPCMS[4,5]<- sum(dffc$dm_kineo_met)
zonesPCMS[4,6]<- sum(dffc$dm_kineo_other)

#Total viatges en un dia laborable dins la Primera Corona
zonesPCMS[2,2]<-sum(dffc[intpc,43:46])
zonesPCMS[2,3]<- sum(dffc$dm_kineo_car[intpc])
zonesPCMS[2,4]<- sum(dffc$dm_kineo_bus[intpc])
zonesPCMS[2,5]<- sum(dffc$dm_kineo_met[intpc])
zonesPCMS[2,6]<- sum(dffc$dm_kineo_other[intpc])

zonesPCMS2[2,2]<-(zonesPCMS[2,3]/zonesPCMS[2,2])*100
zonesPCMS2[2,3]<-(zonesPCMS[2,4]/zonesPCMS[2,2])*100
zonesPCMS2[2,4]<-(zonesPCMS[2,5]/zonesPCMS[2,2])*100
zonesPCMS2[2,5]<-(zonesPCMS[2,6]/zonesPCMS[2,2])*100
#zonesPCMS2[2,6]<-(zonesPCMS[2,2]/zonesPCMS[4,2])*100

#Total viatges en un dia laborable de BCN a PC o visceversa
zonesPCMS[3,2]<-sum(dffc[c(from.bcn2pc,from.pc2bcn),43:46])
zonesPCMS[3,3]<-sum(dffc$dm_kineo_car[c(from.bcn2pc,from.pc2bcn)])
zonesPCMS[3,4]<-sum(dffc$dm_kineo_bus[c(from.bcn2pc,from.pc2bcn)])
zonesPCMS[3,5]<-sum(dffc$dm_kineo_met[c(from.bcn2pc,from.pc2bcn)])
zonesPCMS[3,6]<-sum(dffc$dm_kineo_other[c(from.bcn2pc,from.pc2bcn)])

zonesPCMS2[3,2]<-(zonesPCMS[3,3]/zonesPCMS[3,2])*100
zonesPCMS2[3,3]<-(zonesPCMS[3,4]/zonesPCMS[3,2])*100
zonesPCMS2[3,4]<-(zonesPCMS[3,5]/zonesPCMS[3,2])*100
zonesPCMS2[3,5]<-(zonesPCMS[3,6]/zonesPCMS[3,2])*100
#zonesPCMS2[3,6]<-(zonesPCMS[3,2]/zonesPCMS[4,2])*100

#Total viatges en un dia laborable dins de Barcelona
zonesPCMS[1,2]<-sum(dffc[intbcn,43:46])
zonesPCMS[1,3]<-sum(dffc$dm_kineo_car[intbcn])
zonesPCMS[1,4]<-sum(dffc$dm_kineo_bus[intbcn])
zonesPCMS[1,5]<-sum(dffc$dm_kineo_met[intbcn])
zonesPCMS[1,6]<-sum(dffc$dm_kineo_other[intbcn])

zonesPCMS2[1,2]<-(zonesPCMS[1,3]/zonesPCMS[1,2])*100
zonesPCMS2[1,3]<-(zonesPCMS[1,4]/zonesPCMS[1,2])*100
zonesPCMS2[1,4]<-(zonesPCMS[1,5]/zonesPCMS[1,2])*100
zonesPCMS2[1,5]<-(zonesPCMS[1,6]/zonesPCMS[1,2])*100
#zonesPCMS2[1,6]<-(zonesPCMS[1,2]/zonesPCMS[4,2])*100

pandoc.table(zonesPCMS)
pandoc.table(zonesPCMS2)

zonesPCMS21 <- melt(zonesPCMS2, id.var="Zona")
colnames(zonesPCMS21)<- c('Zona','Transport','value')

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,"grey40"))
colours_pal <- colfunc(nlevels(zonesPCMS21$Transport))
names(colours_pal) <- levels(zonesPCMS21$Transport)

g9<- ggplot(zonesPCMS21,aes(x= as.factor(Zona), y=value, fill=
Transport))

```

```

g9+geom_bar(stat='identity')+scale_fill_manual(values=colours_pal,name
= "")+
theme_classic()+ggtitle("Primera Corona segons mode de transport") +
  theme(axis.line.x = element_blank(),axis.line.y = element_blank(),
        plot.title = element_text(hjust = 0.5,size = 15),legend.text =
element_text(size=12),
        axis.title.x = element_text(size = 15),axis.title.y =
element_text(size = 15))+
  xlab("Zona")+ylab("Viatges (%)")

9<-ggplot(zonesPCMS21, aes(x="",y=value))+
  geom_bar(stat='identity',aes(fill=Transport))
g9 + facet_grid(Zona~Transport)+scale_fill_manual(values=colours_pal)+
theme_bw()+xlab("")+guides(fill=FALSE)+
  ggtitle("Primera Corona - Mode de transport")+theme(plot.title =
element_text(hjust = 0.5))

#####
#####PRIMERA CORONA TRANSPORT MOTORITZAT#####
#####
#S'utilitza la mateixa base de dades de l'apartat anterior
#dffc

zonesPCMSM <-
data.frame(Zona=c('IntraBCN','IntraPC','Commuters','Total'),Viatges=rep
(1,4),Cotxe=rep(1,4),Bus=rep(1,4),Metro=rep(1,4))
zonesPCMSM2 <-
data.frame(Zona=c('IntraBCN','IntraPC','Commuters'),Cotxe=rep(1,3),Bus
=rep(1,3),Metro=rep(1,3))

#Total viatges PC
zonesPCMSM[4,2]<- sum(dfffc[,43:45])
zonesPCMSM[4,3]<- sum(dfffc$dm_kineo_car)
zonesPCMSM[4,4]<- sum(dfffc$dm_kineo_bus)
zonesPCMSM[4,5]<- sum(dfffc$dm_kineo_met)

#Total viatges en un dia laborable dins la Primera Corona
zonesPCMSM[2,2]<-sum(dfffc[intpc,43:45])
zonesPCMSM[2,3]<- sum(dfffc$dm_kineo_car[intpc])
zonesPCMSM[2,4]<- sum(dfffc$dm_kineo_bus[intpc])
zonesPCMSM[2,5]<- sum(dfffc$dm_kineo_met[intpc])

zonesPCMSM2[2,2]<-(zonesPCMSM[2,3]/zonesPCMSM[2,2])*100
zonesPCMSM2[2,3]<-(zonesPCMSM[2,4]/zonesPCMSM[2,2])*100
zonesPCMSM2[2,4]<-(zonesPCMSM[2,5]/zonesPCMSM[2,2])*100

#Total viatges en un dia laborable de BCN a PC o visceversa
zonesPCMSM[3,2]<-sum(dfffc[c(from.bcn2pc,from.pc2bcn),43:45])
zonesPCMSM[3,3]<-sum(dfffc$dm_kineo_car[c(from.bcn2pc,from.pc2bcn)])
zonesPCMSM[3,4]<-sum(dfffc$dm_kineo_bus[c(from.bcn2pc,from.pc2bcn)])
zonesPCMSM[3,5]<-sum(dfffc$dm_kineo_met[c(from.bcn2pc,from.pc2bcn)])

zonesPCMSM2[3,2]<-(zonesPCMSM[3,3]/zonesPCMSM[3,2])*100
zonesPCMSM2[3,3]<-(zonesPCMSM[3,4]/zonesPCMSM[3,2])*100
zonesPCMSM2[3,4]<-(zonesPCMSM[3,5]/zonesPCMSM[3,2])*100
#zonesPCMSM2[3,6]<-(zonesPCMSM[3,2]/zonesPCMSM[4,2])*100
#Total viatges en un dia laborable dins de Barcelona
zonesPCMSM[1,2]<-sum(dfffc[intbcn,43:45])
zonesPCMSM[1,3]<-sum(dfffc$dm_kineo_car[intbcn])
zonesPCMSM[1,4]<-sum(dfffc$dm_kineo_bus[intbcn])

```

```

zonesPCMSM[1,5]<-sum(dfffc$dm_kineo_met[intbcbn])

zonesPCMSM2[1,2]<-(zonesPCMSM[1,3]/zonesPCMSM[1,2])*100
zonesPCMSM2[1,3]<-(zonesPCMSM[1,4]/zonesPCMSM[1,2])*100
zonesPCMSM2[1,4]<-(zonesPCMSM[1,5]/zonesPCMSM[1,2])*100
#zonesPCMSM2[1,6]<-(zonesPCMSM[1,2]/zonesPCMSM[4,2])*100

pandoc.table(zonesPCMSM)
pandoc.table(zonesPCMSM2)

zonesPCMSM21 <- melt(zonesPCMSM2, id.var="Zona")
colnames(zonesPCMSM21)<- c('Zona','Transport','value')

g10<- ggplot(zonesPCMSM21,aes(x= as.factor(Zona), y=value, fill=
Transport))
g10+geom_bar(stat='identity')+scale_fill_manual(values=c('red','darkgr
een','darkblue'),name = "Mode de Transport")+
  theme_classic()+ggtitle('Percentatge de mobilitat motoritzada
Primera Corona (%)') +
  theme(axis.line.x = element_blank(),axis.line.y =
element_blank(),plot.title = element_text(hjust = 0.5))+
xlab("Zona")+ylab("")

amount_it <- 12
colours <- rainbow(amount_it) #set colour code
it<-4
color <- colours[it]
colfunc <- colorRampPalette(c(color,'grey40'))
colours_pal <- colfunc(nlevels(zonesPCMSM21$Transport))
names(colours_pal) <- levels(zonesPCMSM21$Transport)

g10<-ggplot(zonesPCMSM21, aes(x="",y=value))+
  geom_bar(stat='identity',aes(fill=Transport))
g10 +
  facet_grid(Zona~Transport)+scale_fill_manual(values=colours_pal)+
  theme_bw()+xlab("")+
  ylab("Viatges (%)")+guides(fill=FALSE)+ ggtitle("Primera Corona
motoritzada")+
  theme(plot.title = element_text(hjust = 0.5,size = 15), axis.title.y
= element_text(size=15))

#####
#####RATIO PÚBLIC/PRIVAT#####
#####
ratio <- matrix(rep(0,20), ncol=5,nrow=4)
ratio <- data.frame(ratio)
colnames(ratio)<- c('Zona','Públic','Privat','Total','Ratio')
ratio[,1]<- c('IntraBCN','IntraPC','Commuters','Total')

ratio[1,2] <- zonesPCMSM[1,4]+zonesPCMSM[1,5]
ratio[2,2] <- zonesPCMSM[2,4]+zonesPCMSM[2,5]
ratio[3,2] <- zonesPCMSM[3,4]+zonesPCMSM[3,5]
ratio[1,3] <- zonesPCMSM[1,3]
ratio[2,3] <- zonesPCMSM[2,3]
ratio[3,3] <- zonesPCMSM[3,3]
ratio[4,2] <- sum(ratio[,2])
ratio[4,3] <- sum(ratio[,3])
ratio[1,4] <- sum(ratio[1,2:3])
ratio[2,4] <- sum(ratio[2,2:3])
ratio[3,4] <- sum(ratio[3,2:3])
ratio[4,4] <- sum(ratio[4,2:3])

```

```

ratio[1,5] <- round(ratio[1,2]/ratio[1,3],2)
ratio[2,5] <- round(ratio[2,2]/ratio[2,3],2)
ratio[3,5] <- round(ratio[3,2]/ratio[3,3],2)
ratio[4,5] <- round(ratio[4,2]/ratio[4,3],2)
ratio

#####
#####MAPA DE CALOR#####
#####
dffc$munos <- factor(dffc$munos)
dffc$munds <- factor(dffc$munds)

ttpc1<-as.data.frame(xtabs(dffc$tot_kineo~dffc$munos+dffc$munds))
names(ttpc1)<-c("munos","munds","viatges")
ttpc1$viatges<-rescale(log(ttpc1$viatges))
base_size <- 9
p1 <- ggplot(ttpc1, aes(munds, munos)) + geom_tile(aes(fill =

viatges),colour="white") +
  scale_fill_gradient(low = "green", high = "red") +
  theme_grey(base_size = base_size) + labs(x = "", y = "") +
  ggtitle("Mapa de calor: OD Municipis") +
  scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand =
c(0,0)) +
  theme(legend.position = "none",axis.ticks = element_blank(),
        axis.text.x = element_text(size = base_size * 0.8,
                                     angle = 330,
                                     hjust = 0,
                                     colour = "grey50"))

p1

ttpc2<-as.data.frame(xtabs(dffc$dm_kineo_car~dffc$munos+dffc$munds))
names(ttpc2)<-c("munos","munds","viatges")
ttpc2$viatges<-rescale(log(ttpc2$viatges))
base_size <- 9
p2 <- ggplot(ttpc2, aes(munds, munos)) + geom_tile(aes(fill =

viatges),colour="white") +
  scale_fill_gradient(low = "green", high = "red") +
  theme_grey(base_size = base_size) + labs(x = "", y = "") +
  ggtitle("Mapa de calor: OD Municipis - Cotxe") +
  scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand =
c(0,0)) +
  theme(legend.position = "none",axis.ticks = element_blank(),
        axis.text.x = element_text(size = base_size * 0.8,
                                     angle = 330,
                                     hjust = 0,
                                     colour = "grey50"))

p2

ttpc3<-as.data.frame(xtabs(dffc$dm_kineo_bus~dffc$munos+dffc$munds))
names(ttpc3)<-c("munos","munds","viatges")
ttpc3$viatges<-rescale(log(ttpc3$viatges))
base_size <- 9
p3 <- ggplot(ttpc3, aes(munds, munos)) + geom_tile(aes(fill =

viatges),colour="white") +
  scale_fill_gradient(low = "green", high = "red") +
  theme_grey(base_size = base_size) + labs(x = "", y = "") +
  ggtitle("Mapa de calor: OD Municipis - Bus") +

```

```

    scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand =
c(0,0)) +
    theme(legend.position = "none",axis.ticks = element_blank(),
          axis.text.x = element_text(size = base_size * 0.8,
                                     angle = 330,
                                     hjust = 0,
                                     colour = "grey50"))
p3

ttpc4<-as.data.frame(xtabs(dffc$dm_kineo_met~dffc$munos+dffc$munds))
names(ttpc4)<-c("munos","munds","viatges")
ttpc4$viatges<-rescale(log(ttpc4$viatges))
base_size <- 9
p4 <- ggplot(ttpc4, aes(munds, munos)) + geom_tile(aes(fill =

viatges),colour="white") +
  scale_fill_gradient(low = "green", high = "red") +
  theme_grey(base_size = base_size) + labs(x = "", y = "") +
  ggtitle("Mapa de calor: OD Municipis - Metro") +
  scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand =
c(0,0)) +
  theme(legend.position = "none",axis.ticks = element_blank(),
        axis.text.x = element_text(size = base_size * 0.8,
                                   angle = 330,
                                   hjust = 0,
                                   colour = "grey50"))
p4

ttpc5<-as.data.frame(xtabs(dffc$dm_kineo_other~dffc$munos+dffc$munds))
names(ttpc5)<-c("munos","munds","viatges")
ttpc5$viatges<-rescale(log(ttpc5$viatges))
base_size <- 9
p5 <- ggplot(ttpc5, aes(munds, munos)) + geom_tile(aes(fill =

viatges),colour="white") +
  scale_fill_gradient(low = "green", high = "red") +
  theme_grey(base_size = base_size) + labs(x = "", y = "") +
  ggtitle("Mapa de calor: OD Municipis - Altres") +
  scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand =
c(0,0)) +
  theme(legend.position = "none",axis.ticks = element_blank(),
        axis.text.x = element_text(size = base_size * 0.8,
                                   angle = 330,
                                   hjust = 0,
                                   colour = "grey50"))
p5

#####
#####ACP PC#####
#####
vars_num<-c(5:30,37:38,43:46);length(vars_num)
taula1<- dffc[,vars_num]

#Sense les diagonals, on la bdd és de la PC
sel<-which(dffc$ori==dffc$des)
dffc2 <- dffc[-sel,-47]
taula1<- taula1[-sel,]

acp1<-PCA(taula1, quanti.sup = c(1:2,27:32))
plot(acp1,invisible = "ind.sup", title='Projecció 1-2 dels individus')

```

```
summary(acp1)
print(acp1)
acp1$eig
plot(acp1$eig[,1],type="l",main="Regla últim colze")
abline(v=4,col='red')

fviz_cos2(acp1, choice = "var", axes = 1:2)+theme_bw()
fviz_pca_var(acp1, col.var="cos2",repel=TRUE)+
  scale_color_gradient2(low="green", mid="blue",
                        high="red", midpoint=0.75)+theme_bw()

plot.PCA(acp1, axes=c(1, 2), choix="ind")

# Cal determinar els outliers i usar-los com a suplementaris
#####
indcoord <- as.matrix(acp1$ind$coord[,1:2])
mvout<-aq.plot(indcoord[,1:2],seq(0,1,by=0.1))

indcoord2 <- indcoord
indcoord2$mvout<-factor(mvout$outliers)
catdes(indcoord2 ,which(names(indcoord2 )=="mvout"))
#####
l<-boxplot(acp1$ind$coord[,1])
summary(acp1$ind$coord[,1])
calcQ(acp1$ind$coord[,1])
quantile(abs(acp1$ind$coord[,1]),seq(0,1,0.005))
l1<-which(abs(acp1$ind$coord[,1])>6.74);length(l1)
quantile(abs(acp1$ind$coord[,2]),seq(0,1,0.005))
l12<-which(abs(acp1$ind$coord[,2])>1.33);length(l12)
l1 <- unique(l1,l12)
dffc3<- dffc2[-l1,]

taula2<-taula1[-l1,]
acp2<-PCA(taula1, ind.sup = l1,quanti.sup = c(1:2,27:32))
plot(acp2, choix="ind",invisible = "ind.sup",title='Projecció 1-2 dels
individus')
summary(acp2)
acp2$eig

plot(acp2$eig[,1],type="l",main="Regla últim colze",
invisible="ind.sup")
abline(v=4,col='red')

fviz_cos2(acp2, choice = "var", axes = 1:2)+
  ggtitle('Cos2 de les variables als eixos 1-2')+
  ylab('Cos2:Qualitat de la representació')+
  theme(axis.line.x = element_line(c(0:23))
        ,plot.title = element_text(hjust = 0.5,size=15))

fviz_pca_var(acp2, col.var="cos2",repel=TRUE)+
  scale_color_gradient2(low="green", mid="blue",
                        high="red", midpoint=0.25)+
  ggtitle('Projecció de les variables als eixos 1-2')+
  theme(plot.title = element_text(hjust = 0.5,size=15))

fviz_cos2(acp2, choice = "var", axes = 3:4)+
  ggtitle('Cos2 de les variables als eixos 3-4')+
  ylab('Cos2:Qualitat de la representació')+
  theme(axis.line.x = element_line(c(0:23))
        ,plot.title = element_text(hjust = 0.5,size=15))
fviz_pca_var(acp2, col.var="cos2",repel=TRUE, axes = 3:4)+
```

```

scale_color_gradient2(low="green", mid="blue",
                      high="red", midpoint=0.25)+
ggtitle('Projecció de les variables als eixos 3-4')+
theme(plot.title = element_text(hjust = 0.5,size=15))

acp2<-PCA(taula1, ind.sup = 11, quanti.sup = c(1:2,27:32),ncp = 4)
acp2$var$cor

fviz_eig(acp2, addlabels = TRUE, ylim = c(0, 50))

#####
#Projecció de les variables categoriques a l'ACP
#####
Psi=acp2$ind$coord

dffc3$munori<-factor(dffc3$munori)
cenori1 <- tapply(Psi[,1],dffc3$munos, mean)
cenori2 <- tapply(Psi[,2],dffc3$munos, mean)
plot(acp2, choix="ind",invisible = c("ind.sup","ind"))
text(cenori1,cenori2, labels=levels(dffc3$munos),
col=1:length(cenori1), cex=1)

dffc3$munds<-factor(dffc3$munds)
cendes1 <- tapply(Psi[,1],dffc3$munds, mean)
cendes2 <- tapply(Psi[,2],dffc3$munds, mean)
plot(acp2, choix="ind",invisible = c("ind.sup","ind"))
text(cendes1,cendes2, labels=levels(dffc3$munds),
col=1:length(cendes1))

#####
#####ACP PROFILING#####
#####

nd=4
Psi = acp2$ind$coord

set.seed(123)
k1<-NULL
k2<-NULL
for(i in 1:50){
  k1[i] <- kmeans(Psi, centers = i)$betweens #maximitzar
  k2[i] <- kmeans(Psi, centers = i)$tot.withinss #minimitzar
}
plot(1:50, k1, type='b',xlab='Nombre de clústers')
lines(k2, type = 'b')
abline(v=15,col='red')

set.seed(123)
cl <- kmeans(Psi,15)
table(cl$cluster)

plot(Psi[,3], Psi[,4], col=cl$cluster, main = "Agrupació amb 15
classes",pch=20)
#legend('topright', c(1:15), pch=20)
#colors()[1:15]

tapply(dffc3$tot_kineo,cl$cluster, sum)
tapply(dffc3$tot_kineo,cl$cluster, mean)

#Cotxe

```

```

round(tapply(dffc3$dm_kineo_car,cl$cluster, sum),1)
round(tapply(dffc3$dm_kineo_car,cl$cluster, mean),2)

#Bus
round(tapply(dffc3$dm_kineo_bus,cl$cluster, sum),1)
round(tapply(dffc3$dm_kineo_bus,cl$cluster, mean),2)

#Metro
round(tapply(dffc3$dm_kineo_met,cl$cluster, sum),1)
round(tapply(dffc3$dm_kineo_met,cl$cluster, mean),2)

#Altres
round(tapply(dffc3$dm_kineo_other,cl$cluster, sum),1)
round(tapply(dffc3$dm_kineo_other,cl$cluster, mean),2)

#Comarques amb tot_kieno
dffc3$comori<-factor(dffc3$comori)
table(dffc3$comdes,cl$cluster)

round(xtabs(dffc3$tot_kineo~dffc3$comori+cl$cluster),1)

dffc3$comdes<-factor(dffc3$comdes)
round(xtabs(dffc3$tot_kineo~dffc3$comdes+cl$cluster),1)

#Comarques amb car
round(xtabs(dffc3$dm_kineo_car~dffc3$comori+cl$cluster),1)
round(xtabs(dffc3$dm_kineo_car~dffc3$comdes+cl$cluster),1)

#Comarques amb Bus
round(xtabs(dffc3$dm_kineo_bus~dffc3$comori+cl$cluster),1)
round(xtabs(dffc3$dm_kineo_bus~dffc3$comdes+cl$cluster),1)

#Comarques amb Metro
round(xtabs(dffc3$dm_kineo_met~dffc3$comori+cl$cluster),1)
round(xtabs(dffc3$dm_kineo_met~dffc3$comdes+cl$cluster),1)

#Comarques amb Metro
round(xtabs(dffc3$dm_kineo_other~dffc3$comori+cl$cluster),1)
round(xtabs(dffc3$dm_kineo_other~dffc3$comdes+cl$cluster),1)

#####
#####ACS PC#####
#####
#dffc bdd
dffc2<-dffc2

dffc2$munos<-factor(dffc2$munos)
dffc2$munds<-factor(dffc2$munds)

#Total kineo amb municipi
tt<-xtabs(dffc2$tot_kineo~dffc2$munos+dffc2$munds);tt
chisq.test(tt)#p-value < 2.2e-16
sum(dffc2$tot_kineo)
resca<-CA(tt)
summary(resca)

fviz_ca_biplot(resca,axes=1:2,ggtheme = theme_minimal())+
  scale_color_gradient2(low="green", mid="blue",
    high="red", midpoint=0.90)+theme_bw()

```



```

fviz_ca_row(resca, axes=1:2, col.row = "contrib", arrow = c(TRUE, TRUE),
            gradient.cols = c("green", "cyan", "blue"),
            repel = TRUE)+theme_bw()

#Total kineo entre municipis
tt2<-xtabs(dffca$tot_kineo~dffca$munos+dffca$munds);tt2
chisq.test(tt2)#p-value < 2.2e-16
diag(tt2)<-0
resca2<-CA(tt2)
summary(resca2)

fviz_ca_biplot(resca2, axes=1:2, repel=TRUE,
               title="CA-Biplot: Viatges entre municipis Primera
Corona ") +theme_bw()

#Total kineo entre municipis hores punta 7-10
tt3<-
xtabs((dffca$h7_kineo+dffca$h7p8_kineo+dffca$h9_kineo)~dffca$munos+dffca$munds);tt3
diag(tt3)<-0
chisq.test(tt3)#p-value < 2.2e-16
resca3<-CA(tt3)
summary(resca3)

fviz_ca_biplot(resca3, axes=1:2, repel=TRUE,
               title="CA-Biplot: Viatges entre municipis Primera
Corona franja horària 7-10h") +theme_bw()

#Total kineo entre municipis hores punta 13-15
tt4<-
xtabs((dffca$h13_kineo+dffca$h14_kineo)~dffca$munos+dffca$munds);tt4

diag(tt4)<-0
chisq.test(tt4)#p-value < 2.2e-16
resca4<-CA(tt4)
summary(resca4)

fviz_ca_biplot(resca4, axes=1:2, repel=TRUE,
               title="CA-Biplot: Viatges entre municipis Primera
Corona franja horària 13-15h") +theme_bw()

#Total kineo entre municipis hores punta 17-20
tt5<-
xtabs((dffca$h17_kineo+dffca$h18_kineo+dffca$h19_kineo)~dffca$munos+dffca$munds);tt5

diag(tt5)<-0
chisq.test(tt5)#p-value < 2.2e-16
resca5<-CA(tt5)
summary(resca5)

fviz_ca_biplot(resca5, axes=1:2, repel=TRUE,
               title="CA-Biplot: Viatges entre municipis Primera
Corona franja horària 17-20h") +theme_bw()

#Total kineo amb barri nomes bcn
sel <- (which(dffc2$nombarori!="Barri_Gate" &
dffca2$nombarides!="Barri_Gate"))
dffca2 <- dffc2[sel,]

```

```

dffca2$nombarori<-factor(dffca2$nombarori)
dffca2$nombardes<-factor(dffca2$nombardes)
tt6<-xtabs(dffca2$tot_kineo~dffca2$nombarori+dffca2$nombardes);tt6
chisq.test(tt6)#p-value < 2.2e-16

resca6<-CA(tt6)
summary(resca6)

fviz_ca_biplot(resca6,axes=1:2,ggtheme = theme_minimal())+
  scale_color_gradient2(low="green", mid="blue",
                        high="red", midpoint=0.90)+theme_bw()

fviz_ca_row(resca6,axes=1:2,col.row = "contrib",arrow = c(TRUE, TRUE),
            gradient.cols = c("green", "cyan", "blue"),
            repel = TRUE)+theme_bw()

##Mobilitat entre municipis i districtes de Barcelona
sel<-
which((dffca2$ori>699) | (dffca2$des>699) | dffca2$munori=="Barcelona" | dffca2$
mundes!="Barcelona")
dfm<-dffca2[sel,]
dfm$mundisdes<-factor(dfm$mundisdes)
dfm$munos<-factor(dfm$munos)
dfm$mundisori<-factor(dfm$mundisori)
dfm$munds<-factor(dfm$munds)
tt7<-xtabs(dfm$tot_kineo~dfm$munos+dfm$mundisdes)
chisq.test(tt7)#p-value < 2.2e-16

resca7<-CA(tt7)

fviz_ca_biplot(resca7,axes=1:2,col.row = "contrib",title="KINEO-DM:
all modes",arrow = c(FALSE,TRUE),repel=TRUE)+
  scale_color_gradient2(low="blue", mid="cyan",
                        high="red", midpoint=20)+theme_bw()
fviz_ca_row(resca7,axes=1:2,col.row = "contrib",arrow = c(TRUE, TRUE),
            gradient.cols = c("green", "cyan", "blue"),
            repel = T)+theme_bw()
fviz_ca_col(resca7,axes=1:2,col.col = "contrib",arrow = c(FALSE,
FALSE),
            gradient.cols = c("red", "cyan", "darkblue"),
            repel = T)+theme_bw()

```